Basic Spline Wavelet Transform

and Pitch Detection of a Speech Signal

by

Ken Wenchian Lee

Submitted to the Department of Electrical Engineering and Computer Science
in Partial Fulfillment of the Requirements for the Degrees of

Bachelor of Science

and

Master of Science

at the

Massachusetts Institute of Technology

May 1994

Signature of Author .....
Department of Electrical Engineering and Computer Science

Certified by .....
Professor Peter Elias
Department of Electrical Engineering and Computer Science
Thesis Supervisor (Academic)

Certified by .....
Dr. Forrest Tzeng
Company Supervisor (COMSAT Laboratories)

Accepted by .....
Professor Frederic R. Morgenthaler
Chairman, Committee on Graduate Students

Basic Spline Wavelet Transform

and Pitch Detection of a Speech Signal

by

Ken Wenchian Lee

Submitted to the Department of Electrical Engineering and Computer Science
on January 28, 1994, in Partial Fulfillment of the Requirements for the Degrees of

Bachelor of Science

and

Master of Science

Abstract

In recent years the wavelet transform and its applications have gained tremendous interest and have come under intense investigation. Successful applications of the transform are inherently dependent on the wavelets themselves, whose properties necessary for a specific application are not well known. In this thesis, besides examining wavelet theory, we study the wavelet issues important in the accurate pitch detection of a speech signal in no noise, low noise, and high noise environments.

To have control over the properties of the wavelets, we employ a transform based on wavelets built from a sum of shifted B-spline functions. With these designed wavelets, the transform is used to clarify the issues of support, smoothness, and viewing scale and their importance. Performance in pitch detection is judged by how well the non-stationary events of glottal closure are translated into local maxima in the transform coefficients and how well these local maxima coincide with the actual instants of glottal closure.

Of all the experimental designs, the minimally supported, antisymmetric wavelet provides the best performance in a no noise environment; in a low noise environment, the performance is satisfactory; in a high noise environment, performance is satisfactory when the wavelet shape is slightly altered. Performance of the wavelet in multi-speaker speech is also evaluated and is found to be very similar to that in noise.

Thesis Supervisors : Professor Peter Elias
Department of Electrical Engineering and Computer Science

Dr. Forrest Tzeng
Scientist, Voiceband Processing Department
COMSAT Laboratories

Acknowledgment

Very special thanks go to Dr. Tzeng, whose guidance and patience in not only this work but also previous assignments have made my VI-A experience at COMSAT Laboratories enjoyable, memorable, and highly educational. I also would like to thank Professor Elias at MIT and Dr. Unser at NIH for their guidance, comments, and advice. Financial and technical support from COMSAT Laboratories, particularly the Voiceband Processing Department of the Communications Technology Division, is gratefully acknowledged.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

Recently, wavelet analysis and its applications have been under intense investigation in the scientific and engineering communities. A reason for this is that wavelet analysis elegantly unifies the seemingly disparate areas of wavelet series expansions in applied mathematics, subband coding in speech and image processing, and multiresolution signal decomposition in computer vision [6]. In addition, the wavelet transform is potentially more useful in non-stationary signal analysis than traditional Fourier methods.

A convenient way to view the wavelet transform is as a mathematical tool that decomposes a function onto a family of basis functions. But how well the transform fares in any application is inherently dependent on the choice of these functions. The objective of this work is a study of the factors and trade-offs involved in making these choices. The specific application under consideration is pitch detection of a speech signal, and the decision-making process learned from this application can be generalized for others.

Because our application deals with speech signals which are relatively smooth, we focus on wavelets built from smooth basic spline functions. But before we discuss the spline-based wavelet transform, some preliminary background information is presented. The thesis is organized as follows:

- Chapter 2 presents the methods and limitations of classical Fourier analysis, an important precursor of wavelet analysis. The presentation on the specific methods serves to preview the similarities and differences between the two analytical frameworks. The limitations of Fourier analysis serve as a motivation to study and employ wavelet analysis.

- Chapter 3 deals with the wavelet transform: the continuous version, followed by the discrete and dyadic versions.
- Chapter 4 discusses the basic spline functions and the wavelet transform based on them.
- Chapter 5 furnishes the performances and evaluations of the basic spline wavelet transform in various pitch detection experiments.
- Chapter 6 concludes the experiments and the entire work.

Mathematical terminology used throughout the development of wavelet theory can be found in the Appendix A. Proofs are provided in Appendix B and C.

Chapter 2 Fourier Analysis

2.1 Introduction

In this chapter, the basics of classical Fourier analysis are presented so that their similarities and differences with wavelet analysis, which is discussed in the next chapter, will become more apparent. The first part (Sections 2.2 and 2.3) of the chapter is a brief review of the representation of periodic functions with the Fourier series and aperiodic functions with the Fourier transform. In the particular case of non-stationary signal analysis, short-time Fourier methods (Section 2.4) as an improvement over conventional Fourier methods are also included. The very last section of the chapter summarizes the limitations of the short-time Fourier transform in non-stationary signal analysis.

2.2 Fourier Series

The Fourier series is a convenient way to represent a finite-power *periodic* signal $f(t)$ in terms of a linear combination of complex exponentials,

$$f(t) = \sum_{n=-\infty}^{\infty} a_n e^{-int}, \qquad (2.2\text{-}1)$$

where $i = \sqrt{-1}$. In other words, a T-periodic function $f(t)$ can be decomposed in terms of mutually orthogonal basis functions, $b(nt) = e^{int}$, generated from integral dilations of the prototype

$$b(t) = e^{it}, \qquad (2.2\text{-}2)$$

such that

$$\langle b(nt), b(mt) \rangle = \frac{1}{T} \int_0^T e^{int} e^{imt} dt = 0, \quad for\ n \neq m. \tag{2.2-3}$$

Here, $\langle \cdot, \cdot \rangle$ denotes inner product.

Because the basis functions are mutually orthogonal, the series coefficients can be calculated as

$$a_n = \frac{1}{T} \int_0^T f(t) e^{-int} dt. \tag{2.2-4}$$

In the signal processing literature, equations (2.2-1) and (2.2-4) are commonly referred to as the *synthesis* and *analysis* equations, respectively.

To further examine the series representation, let us substitute Euler's relation

$$e^{int} = cos nt + i sin nt \tag{2.2-5}$$

into (2.2-1) yielding

$$f(t) = \sum_{n=-\infty}^{\infty} a_n cos nt \quad + \quad i \sum_{n=-\infty}^{\infty} a_n sin nt, \tag{2.2-6}$$

so that a periodic signal can also be represented as a sum of "sinusoidal waves" of distinct frequencies, $\frac{n}{2\pi}$.

In the case of aperiodic signals, an extension of the Fourier series effected by letting the period of the input signal approach infinity naturally leads to the Fourier transform.

2.3 Fourier Transform

Given a finite-energy aperiodic signal $f(t)$, the Fourier transform (FT) analysis-synthesis pair are defined as

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt, \qquad \text{(analysis)} \tag{2.3-1}$$

and

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\omega) e^{i\omega t} d\omega. \qquad \text{(synthesis)} \tag{2.3-2}$$

The transform maps a continuous-time signal onto the continuous-frequency domain. Signal energy is preserved according to the Parseval Relation,

$$\int_{-\infty}^{\infty} |f(t)|^2 \, dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| \hat{f}(\omega) \right|^2 d\omega. \qquad (2.3\text{-}3)$$

However, in practice, the Fourier transform - also known as the natural "stationary transform"[6] - is rarely used. From the FT analysis equation, we see that the Fourier coefficient at frequency ω tells us the prevalence of this frequency in the entire signal so that an integration over all time is required. When Euler's relation is substituted into (2.3-2), we see that an aperiodic signal is represented by sine waves localized in frequency but global in time. Such a representation is adequate only if the signal is made up of distinct stationary components. But such signals rarely exist. Furthermore, because any sudden change in time is reflected in the entire spectrum, the Fourier transform is inappropriate for signal analysis because it does not provide any information on the time location of interesting spectral components.

The inadequacies of the FT can be summarized as

- Local changes in time are reflected in the entire frequency domain.
- Lack of information on the time location of spectral components.

An extensively studied method devised to compensate somewhat for these inadequacies is the short-time Fourier transform, which is the topic of discussion of the next section. Chapter 3 discusses a better overall solution to these fundamental problems.

2.4 Short-Time Fourier Analysis

2.4.1 Signal Analysis

Instead of working with the entire signal, short-time Fourier analysis deals with a windowed portion of it. For a window $w(t)$, the short-time Fourier transform (STFT) of $f(t)$ is defined as

$$\hat{f}_{ST}(\tau, \omega) = \int_{-\infty}^{\infty} f(t) w(t-\tau) e^{-i\omega t}\, dt. \qquad \text{(analysis)} \qquad (2.4.1\text{-}1)$$

In the case of a discrete-time signal $f(n)$ and a discrete-time window $w(n)$, the STFT becomes

$$\hat{f}_{ST}(n, \omega) = \sum_{m=-\infty}^{\infty} f(m) w(n-m) e^{-i\omega m}, \quad \text{(analysis-discrete)} \quad (2.4.1\text{-}2)$$

which says that the entire signal is windowed (Fig. 2.4.1-1) before the conventional Fourier transform is taken.

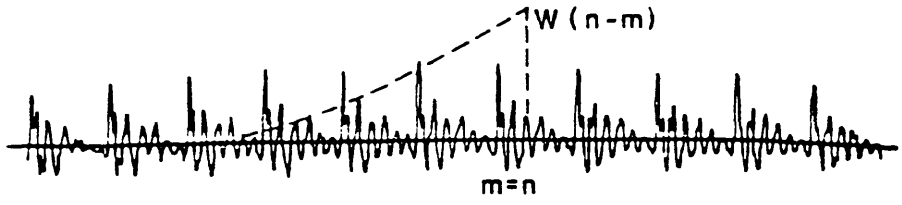


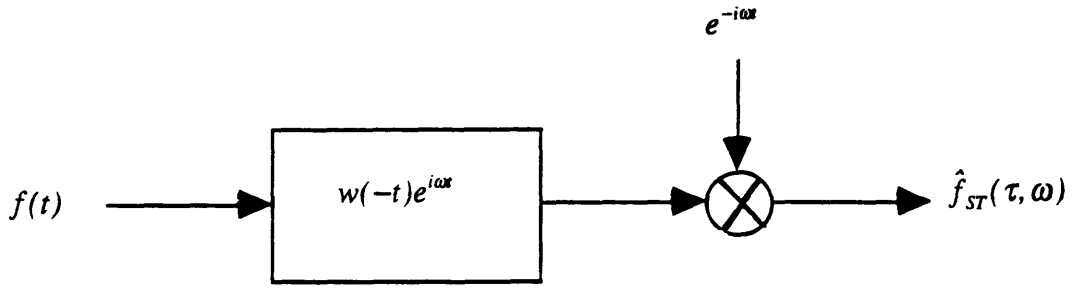
Figure 2.4.1-1: Windowing operation of the STFT. (After Allen and Rabiner [2]).

While the FT furnishes an infinite-time spectrum of a signal, the STFT provides a short-time spectrum. And in contrast to the FT's need for signal stationarity, the STFT requires *quasi-stationarity*, i.e., the signal must be stationary within the time interval of interest. Hence, like a musical score, the STFT provides some information on the dependence of frequency on time [6].

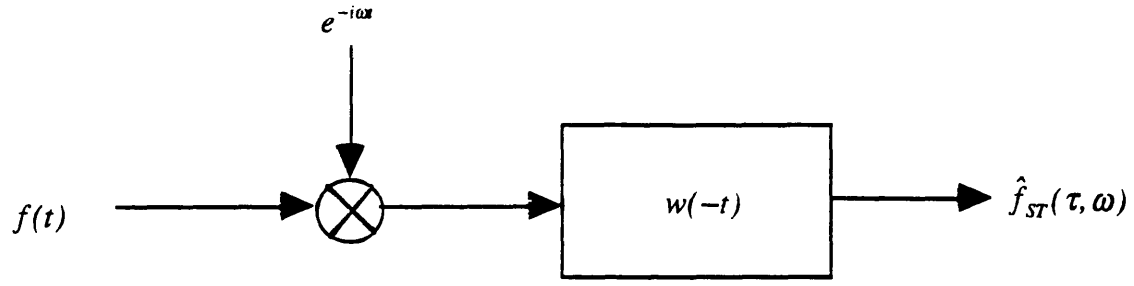
The STFT analysis equation can also be viewed from a filter bank perspective. Rewriting the equation in terms of the Fourier transforms of the signal and window, $\hat{f}(\omega)$ and $\hat{w}(\omega)$, respectively, gives

$$\hat{f}_{ST}(\tau,\omega) = e^{-i\omega\tau}\int_{-\infty}^{\infty}\hat{f}(\theta)\hat{w}(\theta-\omega)e^{i2\pi\theta\tau}\,d\theta. \qquad (2.4.1\text{-}3)$$

Thus, the coefficients can also be obtained by filtering $f(t)$ with a bandpass filter and then demodulating the result back to the origin of the frequency axis (Fig. 2.4.1-2a). And if the order of the filtering and demodulation operations is reversed, a lowpass implementation results (Fig. 2.4.1-2b).



(a) Bandpass Filter



(b) Lowpass Filter

Figure 2.4.1-2: STFT via filter banks.
(a) Implementing STFT with a bandpass filter.
(b) Implementing STFT with a lowpass filter.

Both the windowing and the filter bank interpretations of the STFT are illustrated in Fig. 2.4.1-3b. The horizontal axis corresponds to the time window interval, and the vertical axis to the filter bandwidth.

To examine the effectiveness of the STFT in signal analysis, we need to study its time-frequency resolution capabilities. Let a window and its Fourier transform be represented as $w(t)$ and $\hat{w}(\omega)$, respectively. The window's frequency bandwidth, $\Delta\omega$, is defined as

$$\Delta\omega = \sqrt{\frac{\int_{-\infty}^{\infty} \omega^2 |\hat{w}(\omega)|^2 \, d\omega}{\int_{-\infty}^{\infty} |\hat{w}(\omega)|^2 \, d\omega}}. \qquad (2.4.1\text{-}4)$$

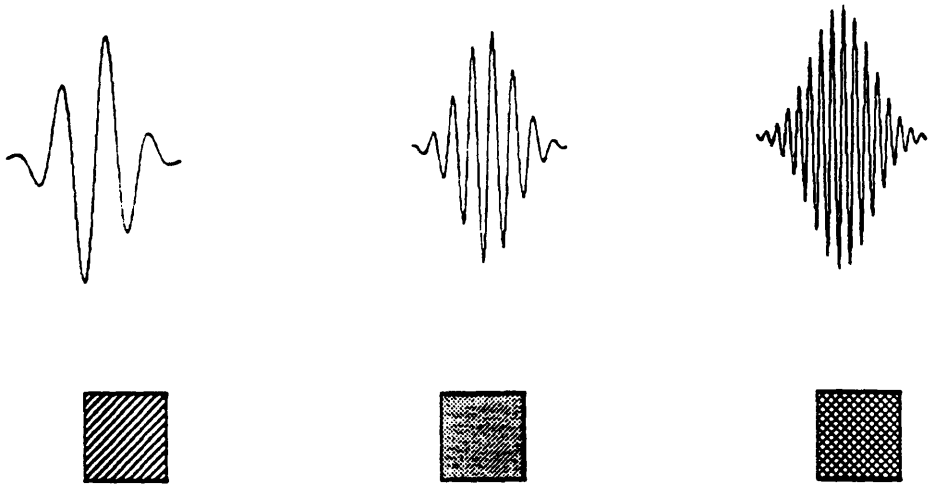
Analogously, its time width Δt is

$$\Delta t = \sqrt{\frac{\int_{-\infty}^{\infty} t^2 |w(t)|^2 \, dt}{\int_{-\infty}^{\infty} |w(t)|^2 \, dt}}. \qquad (2.4.1\text{-}5)$$

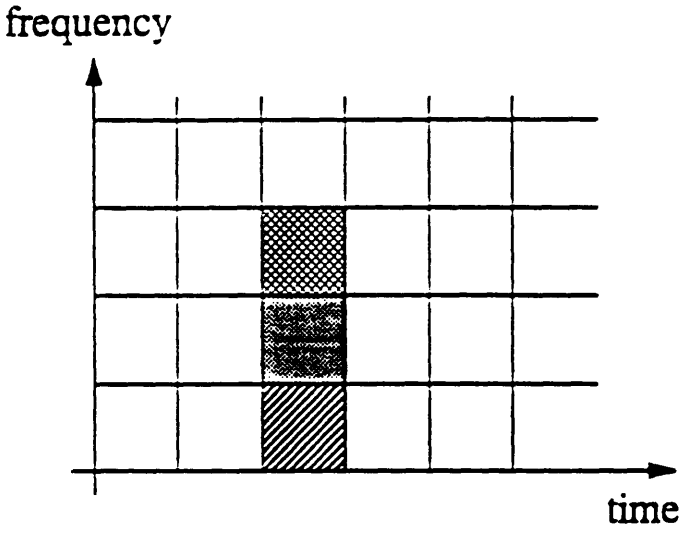
For good time resolution, Δt should be as small as possible; for good frequency resolution, $\Delta\omega$ should also be as small as possible. But since they must obey the uncertainty principle[8],

$$\Delta t \, \Delta\omega \geq \frac{1}{2}, \qquad (2.4.1\text{-}6)$$

a time-frequency resolution trade-off always exists. For maximum resolution, equality in (2.4.1-6) must hold, which happens only when the window $w(t)$ is Gaussian.



(a)



(b)

Figure 2.4.1-3 : STFT basis functions; windowing and filter bank interpretations.
(a) STFT basis functions.
(b) Time-frequency plane illustrating both windowing and filter bank views of the STFT.
(After Vetterli and Herley [36]).

2.4.2 Signal Expansion

The counterpart of the STFT analysis equation is the synthesis equation, or inverse short-time Fourier transform, which for a synthesis window $g(t)$ is defined as ([4], [7])

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}_{ST}(\tau, \omega)[g(t-\tau)e^{i\omega t}]d\tau d\omega, \quad \text{(synthesis)} \quad (2.4.2\text{-}1)$$

where $\tau,\ \omega \in \Re$. Unlike the FT's basis functions which are infinite-duration complex exponentials, those of the STFT have finite duration (Fig. 2.4.1-3a).

Hlawatsch and Boudreaux-Bartels have interpreted the window $g(t)$ as centered at $(\tau = 0,\ \omega = 0)$ in the time-frequency plane (Fig. 2.4.1-3b) [7]. The coefficient $\hat{f}_{ST}(\tau, \omega)$ then corresponds to the contribution of the time-frequency point $(\tau = 0,\ \omega = 0)$ to the decomposition of $f(t)$. Each rectangle in the time-frequency plane (Fig. 2.4.1-3b) corresponds to a resolution cell for signal representation. For exact reconstruction, the analysis and synthesis windows must satisfy [4]

$$\int_{-\infty}^{\infty} g(t)w(t)dt = 1. \quad (2.4.2\text{-}2)$$

So far we have considered an expansion of a signal in terms of a continuous time-frequency plane, but a discrete time-frequency plane can also be used. The discretized versions of the STFT analysis and synthesis equations, (2.4.1-1) and (2.4.2-1), respectively, are ([4], [7])

$$\hat{f}(a\tau, b\omega) = \int_{-\infty}^{\infty} f(t)w(t-a\tau)e^{-ib\omega t}dt \quad \text{(discrete-analysis)} \quad (2.4.2\text{-}3)$$

and

$$f(t) = \sum_{a=-\infty}^{\infty} \sum_{b=-\infty}^{\infty} \hat{f}_{ST}(a\tau, b\omega)[g(t-a\tau)e^{ib\omega t}]. \quad \text{(discrete-synthesis)} \quad (2.4.2\text{-}4)$$

where $a, b \in Z$. The condition required to ensure exact reconstruction becomes [7]

$$\frac{2\pi}{\omega} \sum_{a=-\infty}^{\infty} g(t + \frac{2\pi b}{\omega} - a\tau)w(t-a\tau) = \delta[b], \quad t \in \Re. \quad (2.4.2\text{-}5)$$

An interesting point to make here is that the signal expansion of equation (2.4.2-4) can be generalized as

$$f(t) = \sum_{a=-\infty}^{\infty} \sum_{b=-\infty}^{\infty} c_{a,b}\, h_{a,b}(t), \qquad (2.4.2\text{-}6)$$

where $h_{a,b}(t) = h(t - a\tau)e^{ib\omega t}$, $a, b \in Z$. Equation (2.4.2-6) is commonly known as the Gabor expansion of a continuous-time signal into shifts of the window $h(t)$ and scaled versions of the Fourier kernel $e^{i\omega t}$. Maximum time and frequency localization occurs only when $h(t)$ is Gaussian. And if the window in the STFT is Gaussian, the resulting transform is also known as the Gabor transform.

2.4.3 Limitations of Short-Time Fourier Analysis

Though short-time Fourier methods are an improvement over conventional Fourier methods for non-stationary signal analysis, they are still very inadequate. The efficacy of the STFT depends on the assumption that an input signal is stationary within the window of interest so that any non-stationarities within this window will only affect the short-time spectrum. Unlike the Fourier transform, in which any sudden change in time is reflected throughout the spectrum, the STFT has, in essence, localized the spectral disturbances. Hence, the effectiveness of this approach depends critically on the window's time width.

According to the definition of the STFT (equation (2.4.1-1)), only shifts of the window are used for analysis so that two pulses less than a window time width apart cannot be distinguished. Such inflexibility renders the STFT inappropriate for many signal analysis problems, e.g., pitch detection of a speech signal in which pitch pulses vary from 1.25 ms to 40 ms apart [85].

The limitations of the STFT can also be studied from a signal expansion perspective: that is, to study a signal, we decompose it with respect to a set of basis functions and hope the interesting parts of the signal will be more evident in this

decomposed form. In contrast to the infinite-duration basis functions of the Fourier transform, those of the STFT are finite. However, this duration is *constant* (Fig. 2.4.1-3a). Unless we know how fast the signal is changing, the STFT is highly inflexible - and perhaps even inappropriate. If we assume the signal is extremely fast-changing, and therefore, choose a window with an extremely small time width so that we do not lose any detail, the amount of computation involved will render the method impractical.

In summary, the constancy in both window width and basis function duration of the STFT makes it an unattractive tool for non-stationary signal analysis. The wavelet transform, discussed in the next chapter, is a generalization of the Gabor expansion, and by relaxing the constancy restrictions of the STFT, it is more attractive - and potentially more useful - in non-stationary signal analysis, although it still must obey the uncertainty principle.

Chapter 3 Wavelet Analysis

3.1 Introduction

While the basics of the Fourier methods are covered in Chapter 2, this chapter deals with the theoretical and interpretational aspects of the wavelet methods. Section 3.2 provides a brief historical background. Section 3.3 discusses various kinds of wavelet transforms: continuous, discrete, and dyadic. Also included are their properties and interpretations from multiresolution and filter bank perspectives. Section 3.4 delineates the differences in signal analysis capabilities between the Fourier and wavelet methods. The last section discusses some of the reasons for the great interest in the wavelet transform as a signal processing tool, especially in the areas of image and speech processing.

Note that some of the mathematical terminology used throughout this chapter are defined in Appendix A. Proofs are provided in Appendix B and C.

3.2 Historical Background

Although explicit wavelet analysis did not come into play until a decade or so ago, something very similar has existed since 1946 when Gabor proposed a method to study a signal more carefully by shifting an analyzing function along the time and frequency axes. The most popular function used in such a capacity is the Gaussian, in which case the Gabor transform results. As far as signal analysis is concerned, the Gabor transform is better than

the STFT because maximum time-frequency localization prevails. However, the Gabor transform still has the same fundamental limitations as the STFT.

About a decade ago a fundamentally better transform was introduced by Grossman and Morlet [26]. Their method deals with the decomposition of functions into "elementary wavelets," which arise from dilations and translations of a single basic wavelet. Early applications in seismic analysis have found it to be numerically stable. Unlike the Gabor transform which may cause numerical instability when a very high frequency transient signal is reconstructed from the Gaussian basis functions, the newly introduced transform does not possess this liability. The mathematical foundations presented in [26] has spurred theoretical studies and novel applications.

Daubechies *et al.* [25] investigates the transform from an applied mathematical point of view: that is , the expansion of a function in terms of a chosen set of basis functions. If the set constitutes an orthonormal basis, the functional expansion in terms of a linear combination of these functions converges. But to achieve orthonormality, the functions themselves sacrifice *smoothness* (i.e., number of continuouse derivatives), or *regularity*. If the function to be expanded happens to be smooth, the basis will most likely produce extraneous transform coefficients and may introduce false high-frequency components. Non-orthogonal, or quasi-orthogonal, basis functions have also been studied and they have been found to be very useful in some areas of signal analysis, in particular edge detection [35], zero-crossing determination [62], ventricular delayed potentials in the electrocardiogram [32], and seizure detection in the electroencephalogram [57].

Though very important, the issues and trade-offs among the properties of orthogonality, functional support, regularity, and complexity in the basis functions have not been very well studied. For a specific application, a carefully considered compromise among the properties needs to be made. The following chapters of this thesis attempt to elucidate what such a compromise entails.

3.3 Wavelet Transform

3.3.1 Continuous Wavelet Transform

While the basis functions of the Fourier series and Fourier transform are infinite in duration and those of the STFT are windowed versions of the same kind of functions, we would like to do better. We would like the functions themselves to be "small waves," or *wavelets* [12], which decay rapidly to zero, and thereby obviate any use of windows. In addition, these functions must span $L^2(\Re)$ space. And like the complex exponentials, they must have frequencies. In summary, the functions should possess:

- Fast decay.
- A span of the $L^2(\Re)$ space.
- Frequency.

To construct such a basis, we first need a starting function which we can manipulate to create a family of functions. To obtain fast decay, one has to select a starting function that has this property. For the basis to span the entire $L^2(\Re)$ space, one can shift the original function so that the support of the basis covers the entire real axis. To obtain all frequencies, we can contract and dilate the original. For example, for a function $f(t)$ with a domain of $[0, 1]$, the contracted version, $f(2t)$, has twice as many variations within the same domain, while the dilated version, $f(t/2)$, has half as many.

Let the original wavelet be $\psi(t) \in L^2(\Re)$, also called the *mother wavelet*. The constructed basis becomes

$$\psi_{a,b}(t) = \psi(\frac{t-b}{a}), \quad a, b \in \Re. \qquad (3.3.1\text{-}1)$$

Parameter b shifts, or translates, the motion wavelet along the time axis; parameter a scales the rapidity, which is the same kind of scale used in road maps and serves as an analogue of frequency ω in Fourier analysis. For a function $f(t) \in L^2(\Re)$ and a basis $\psi_{a,b}(t) \in L^2(\Re)$, the continuous wavelet transform (CWT) is defined as

$$(CWT_t\, f)(b,a) \;=\; |a|^{-1/2} \int_{-\infty}^{\infty} f(t)\, \overline{\psi(\frac{t-b}{a})}\, dt, \qquad (3.3.1\text{-}2)$$

where $|a|^{-1/2}$ is a normalization factor, and the overhead bar denotes complex conjugation. In terms of an inner product, the definition can be expressed as

$$(CWT_t\, f)(b,a) \;=\; |a|^{-1/2} \langle f(t),\, \psi_{a,b}(t) \rangle. \qquad (3.3.1\text{-}3)$$

The CWT can also be expressed in terms of Fourier transforms:

$$(CWT_\omega\, f)(b,a) \;=\; \frac{|a|^{-1/2}}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\omega)\, e^{ib\omega}\, \overline{\hat{\psi}(a\omega)}\, d\omega, \qquad (3.3.1\text{-}4)$$

where $i = \sqrt{-1}$.

Not every function in $L^2(\Re)$ can be a wavelet. The necessary conditions are:

$$\psi(t) \;\in\; L^1(\Re) \qquad (3.3.1\text{-}5)$$

and

$$\int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|}\, d\omega \;<\; \infty. \qquad (3.3.1\text{-}6)$$

The latter condition is commonly known as the *admissibility condition*, and a function that obeys this condition is said to be *admissible*. Let us see where this condition originates.

Since the basis is built from a mother wavelet, it must be the case that the constructed family of functions can be projected back onto this "mother" function. Working in Hardy space (see Appendix A), $H^2(\Re)$, Grossman and Morlet [26] formulate this requirement mathematically as

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left| \langle e^{u/2}\, \psi(e^u t - v),\; \psi(t) \rangle \right|^2 du\, dv \;<\; \infty, \qquad (3.3.1\text{-}7)$$

where $u, v \in \Re$, and $1/a = e^u$ and $b = \frac{v}{e^u}$ are substituted into the CWT definition. The derivation of the admissibility condition from equation (3.3.1-7) can be found in Appendix B. In practical terms, for a function $\psi(t)$ to qualify as a wavelet,

$$\hat{\psi}(\omega=0)=0,$$

i.e., the function must have an average of zero:

$$\int_{-\infty}^{\infty}\psi(t)dt \;=\; 0. \qquad (3.3.1\text{-}8)$$

Now to see how the CWT can be used to process signals, we look into some of its properties:

- Linearity: CWT is linear.

  Proof: Let $f(t)=\sum_{n=1}^{N} f_n(t)$, where $n, N \in Z, \quad N>1$, and $f_n(t)\in L^2(\Re)$, then

  $$(CWT_t f)(b,a) \;=\; |a|^{-1/2}\int_{-\infty}^{\infty}\sum_{n=1}^{N} f_n(t)\;\overline{\psi(\frac{t-b}{a})}dt.$$

  Interchanging the two linear operations, namely integration and summation, gives

  $$(CWT_t f)(b,a) \;=\; \sum_{n=1}^{N}|a|^{-1/2}\int_{-\infty}^{\infty} f_n(t)\,\overline{\psi(\frac{t-b}{a})}dt.$$

  According to the CWT definition,

  $$(CWT_t f)(b,a) \;=\; \sum_{n=1}^{N}(CWT_t f_n)(b,a)\,.$$

- Shift in time: CWT is time-shift invariant.

  Proof: Let $f_T(t)=f(t-T)\in L^2(\Re)$, where $f(t)\in L^2(\Re)$ and $T\in\Re$, then

  $$(CWT_t f_T)(b,a) \;=\; |a|^{-1/2}\int_{-\infty}^{\infty} f(t-T)\,\overline{\psi(\frac{t-b}{a})}dt.$$

  Let $v = t - T$. Substituting

  $$dt = dv$$

  and

  $$t - b = v - (b - T)$$

  into $(CWT_t f_T)(b,a)$ yields

  $$(CWT_v f_T)(b,a) \;=\; |a|^{-1/2}\int_{-\infty}^{\infty} f(v)\,\overline{\psi(\frac{v-(b-T)}{a})}dv,$$

  or $$(CWT_t f_T)(b,a) \;=\; (CWT_t f)(b-T,a).$$

- Scale change: CWT is scale-change invariant.

  Proof: Let $f_c(t)=|c|^{1/2} f(ct)\in L^2(\Re)$, where $f(t)\in L^2(\Re), \quad c\in\Re$, and $c>0$, then

$$(CWT_t f_c)(b,a) = |a|^{-1/2} \int_{-\infty}^{\infty} |c|^{1/2} f(ct) \overline{\psi(\frac{t-b}{a})} dt.$$

Let $v = ct$. Substituting

$$dt = \frac{1}{c} dv$$

and

$$\frac{t-b}{a} = \frac{v-cb}{ca}$$

into the above equation, we obtain

$$\begin{aligned}(CWT_v f_c)(b,a) &= |a|^{-1/2} \int_{-\infty}^{\infty} |c|^{1/2} f(v) \overline{\psi(\frac{v-cb}{ca})} \frac{1}{c} dv \\ &= (CWT_v f)(cb, ca).\end{aligned}$$

- Energy preservation: The CWT preserves signal energy. Specifically, if $f(t) \in L^2(\Re)$ and $\psi(t)$ is admissible, then

$$\frac{1}{c_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |(CWT_t f)(b,a)|^2 \frac{1}{a^2} da\, db = \int_{-\infty}^{\infty} |f(t)|^2 dt, \qquad (3.3.1\text{-}9)$$

where

$$c_\psi = \int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega. \qquad (3.3.1\text{-}10)$$

A proof can be found in Appendix C.

- Inverse transform: For $f(t) \in L^2(\Re)$ and $(CWT_t f)(b,a) \in L^2(\Re)$, the inverse CWT is defined as

$$f(t) = \frac{1}{c_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |a|^{-1/2} \psi(\frac{t-b}{a}) (CWT_t f)(b,a) \frac{1}{a^2} da\, db, \qquad (3.3.1\text{-}11)$$

where c_ψ is defined in (3.3.1-10). A proof can be found in [26].

Just by examining the definition, we see that the CWT can be a very flexible and useful signal analysis tool. There is much freedom in choosing, or designing, them for a specific application. An early application of the CWT was in transient signal analysis, in which Tuteur [32] used different dilation parameters, $1/a$, to pinpoint abnormalities in an electrocardiogram. The distinct parameters chosen were; $1/a$ = 11, 16, and 22. Hence, although the CWT is very useful, in most cases, the transform is not taken over all dilation

parameters because, in practice, a small number of parameters suffices. When the CWT is taken over all continuous parametric values, it is highly redundant; besides, such an undertaking is extremely computationally burdensome. In short, to be of practical use, the transform must be simplified.

A natural way to simplify the CWT is to discretize the dilation and translation parameters, a and b, which leads to the discrete wavelet transform (DWT).

3.3.2 Discrete Wavelet Transform

To discretize the parameters a and b in the CWT definition, let us choose $a = M^j$ and $b = kM^jN$, where $M, N \in \Re$ and $j, k \in Z$. Expressing the CWT in terms of j and k results in the definition of the discrete wavelet transform (DWT):

$$(DWT_t f)(k,j) = M^{-j/2} \int_{-\infty}^{\infty} f(t)\,\overline{\psi(M^{-j}t - kN)}\,dt. \qquad (3.3.2\text{-}1)$$

An analogous formulation in terms of Fourier transforms is

$$(DWT_\omega f)(k,j) = M^{j/2} \int_{-\infty}^{\infty} \hat{f}(\omega)\,\overline{\hat{\psi}(M^j\omega)}\,e^{ikM^jN\omega}\,d\omega. \qquad (3.3.2\text{-}2)$$

Here, the wavelet basis functions are $M^{-j/2}\,\psi(M^{-j}t - kN)$. While the CWT maps a signal onto a continuous a-b plane, the DWT projects it onto a discrete j-k grid.

Some of the properties of the DWT are very similar to those of the CWT, and so their proofs are omitted in the following list:

- Linearity: The DWT is linear.
- Energy preservation: The DWT preserves energy.
- Inverse transform: An original signal $f(t) \in L^2(\Re)$ can be reconstructed from its DWT coefficients as

$$f(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} (DWT_t f)(k,j)\,M^{-j/2}\,\psi(M^{-j}t - kN). \qquad (3.3.2\text{-}3)$$

However, the following two properties of the DWT differ from those of the CWT:

- Shift in time: The DWT is time-shift *variant.*

  Proof: Let $f_T(t) = f(t-T)$, where $f(t) \in L^2(\Re)$ and $T \in \Re$, then

  $$(DWT_t f_T)(k,j) = M^{-j/2} \int_{-\infty}^{\infty} f(t-T)\,\overline{\psi(M^{-j}t - kN)}\,dt. \quad (3.3.2\text{-}4)$$

  Let $v = t - T$. Substituting

  $$dt = dv$$

  and

  $$t = v + T$$

  into (3.3.2-4) yields

  $$(DWT_v f_T)(k,j) = M^{-j/2} \int_{-\infty}^{\infty} f(v)\,\overline{\psi(M^{-j}v - (k - \frac{M^{-j}T}{N})N)}\,dv. \quad (3.3.2\text{-}5)$$

  Since in general, $\frac{M^{-j}}{N}T \neq T$, then

  $$(DWT_t f_T)(k,j) \neq (DWT_t f)(k-T, j). \quad (3.3.2\text{-}6)$$

- Scale change: DWT is scale-change *variant.*

  Proof: Let $f_c(t) = |c|^{1/2} f(ct) \in L^2(\Re)$, where $f(t) \in L^2(\Re)$, $c \in \Re$, and $c > 0$, then

  $$(DWT_t f_c)(k,j) = M^{-j/2} \int_{-\infty}^{\infty} |c|^{1/2} f(ct)\,\overline{\psi(M^{-j}t - kN)}\,dt. \quad (3.3.2\text{-}7)$$

  Let $v = ct$. Rewriting (3.3.2-7) in terms of v gives

  $$(DWT_v f_c)(k,j) = M^{-j/2} \int_{-\infty}^{\infty} |c|^{1/2} f(v)\,\overline{\psi(M^{-j}\frac{v}{c} - kN)}\frac{dv}{c}. \quad (3.3.2\text{-}8)$$

  Further simplification leads to

  $$(DWT_t f_c)(k,j) \neq (DWT_t f)(ck, cj). \quad (3.3.2\text{-}9)$$

The lack of time-shift invariance of the DWT means that for the transform to be meaningful, the time location must be specified.

3.3.3 Dyadic Wavelet Transform

Interesting interpretations and simple implementation arise when we choose $M = \frac{1}{2}$ and $N = 1$ in the definition of the discrete wavelet transform, in which case the dyadic wavelet transform results:

$$(WT_{2^j} f)(k, j) = 2^{j/2} \int_{-\infty}^{\infty} f(t) \overline{\psi(2^j t - k)} dt, \qquad (3.3.3\text{-}1)$$

where $j, k \in Z$. Here, all scales, (2^j), and all shifts, (2^{-j}), are dyadic. Unlike the DWT, the dyadic wavelet transform has properties which are very similar to those of the CWT. In particular, it is time-shift *invariant* and scale-change *invariant* (Proofs of these properties are very similar to those of the CWT). But unlike the CWT, it is much less redundant.

The basis functions of the WT_{2^j} are $2^{j/2} \psi(2^j t - k)$. An interpretation of these functions from a multiresolution signal decomposition point of view and their relationship with filter banks are the topics of discussion of the following sections.

3.3.4 Multiresolution Analysis, Dyadic Wavelet Basis, and Dyadic Wavelet Transform

A very interesting interpretation of the dyadic wavelet basis can be obtained from multiresolution analysis. To fully appreciate this point of view, let us review a signal processing technique that is often viewed as an important precursor of multiresolution analysis.

The theory of multiresolution analysis, currently very popular in computer vision and image processing, is built upon the Laplacian pyramid algorithm introduced by Burt

and Adelson [60] in 1983. Since images tend to have high pixel-to-pixel correlation, the main goal of the pyramidal scheme is to remove as much correlation as possible to effect a net data compression. To simplify our review, let us assume that the original signal, S_0, is one-dimensional. The algorithm starts off by low-pass filtering S_0. If $L[\bullet]$ denotes the low-pass filtering operation and L_0 denotes the low-pass filtered, or blurred, version of the signal, then

$$L_0 = L[S_0],$$

and the residual component, R_0, can be represented as

$$R_0 = S_0 - L_0.$$

Now if we again apply the filtering operation to the already low-pass filtered signal L_0, we have

$$L_1 = L[L_0],$$

and another residual signal results,

$$R_1 = L_0 - L_1.$$

The alternating filtering and subtraction operations can be continued until the signal is decomposed satisfactorily, in which case we end up with the residual components, $R_0, R_1, \ldots, R_n$, and a most blurred remnant of the original signal, L_n. Fig. (3.3.4-1) summarizes the decomposition stage of the pyramidal algorithm. To reconstruct a signal, a complementary technique is used; Fig. (3.3.4-2) illustrates the reconstruction stage of the pyramidal algorithm.

$$\begin{array}{ccccccccc} S_0 & \rightarrow & L_0 & \rightarrow & L_1 & \rightarrow & \cdots & \rightarrow & L_n \\ \downarrow & & \downarrow & & \downarrow & & & \downarrow & \\ R_0 & & R_1 & & R_2 & & & R_n & \end{array}$$

Figure 3.3.4-1: Decomposition of a signal into lower resolution components.

$$\begin{array}{ccccccccc} S_0 & \leftarrow & L_0 & \leftarrow & \cdots & \leftarrow & L_{n-1} & \leftarrow & L_n \\ \uparrow & & \uparrow & & & & \uparrow & & \\ R_0 & & R_1 & & & & R_n & & \end{array}$$

Figure 3.3.4-2: Reconstruction of higher resolution components of a signal.

Because the residual components, $R_n's$, and the most blurred component, L_n, are decorrelated, only they will be encoded.

The idea of representing a signal at different resolutions (e.g., an original image as an order hierarchy of blurred images) serves as the foundation of Mallat's multiresolution approximation framework ([24], [27], [33]), which involves approximating a function, $f(t) \in L^2(\Re)$, at different resolutions and is described in the rest of this section.

Let us form a basis made up of shifts of a certain function *s(t)* along the time axis

$$s_k(t) = s(t-k), \quad k \in Z,$$

such that the basis spans a subspace of $L^2(\Re)$, V_0. In other words,

$$V_0 = span\{s_k(t)\}.$$

Employed in such a capacity, *s(t)* is commonly known as the *scaling function*. Let us now generate a family of functions by scaling the set of shifted spanning functions, or

$$s_{j,k}(t) = 2^{j/2} s(2^j t - k), \quad j, k \in Z. \tag{3.3.4-1}$$

For a given *j*, a certain subspace V_j is spanned by $s_{j,k}(t)$. As *j* increases the size of the spanned subspace also increases, and an ordered hierarchy of bigger and bigger subspaces can be built:

$$\ldots \subset V_{-1} \subset V_0 \subset V_1 \subset \ldots \subset V_j \tag{3.3.4-2}$$

such that

$$V_j \subset V_{j+1}. \tag{3.3.4-3}$$

In addition, as $j \to \infty$, the entire $L^2(\Re)$ space is spanned; analogously, as $j \to -\infty$ the null space is spanned:

$$V_\infty = L^2(\Re),$$

$$V_{-\infty} = \{0\}.$$

The nesting of subspaces (implied in (3.3.4-3)) means that for any function $f(t) \in L^2(\Re)$,

$$f(t) \in V_j \quad \Leftrightarrow \quad f(2t) \in V_{j+1}. \tag{3.3.4-4}$$

Now that we have seen that bigger and bigger subspaces can be spanned by scaling functions, let us investigate how the residual information between two consecutive subspaces can be represented. Since this residual information is very similar to that briefly discussed in the Laplacian pyramidal algorithm, it can be represented in a similar way. Hence, let W_j denote a subspace that is the orthogonal complement of V_j, or

$$W_j \oplus V_j = V_{j+1}. \tag{3.3.4-5}$$

A projection of a signal onto W_j yields a representation of the residual information. Analogous to the generation of the basis that spans V_j, a basis built from dilations and translations of a wavelet, $\psi(t)$, can be used to span W_j. The resulting basis is

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k), \quad j, k \in Z, \tag{3.3.4-6}$$

which is same as that used in the definition of the dyadic wavelet transform (equation (3.3.3-1)).

From the nesting of the V_j subspaces and the orthogonal complementarity between the subspaces W_j and V_j, an important result arises. By repeatedly substituting W_j subspaces of lower resolution for the V_{j+1} subspaces, we see that the entire $L^2(\Re)$ space can be spanned by a sum of W_j subspaces, or

$$\ldots \oplus W_{j-1} \oplus W_j \oplus W_{j+1} \oplus \ldots = L^2. \tag{3.3.4-7}$$

Consequently, when an arbitrary function needs to be analyzed, the wavelet functions will suffice.

3.3.5 Dyadic Wavelet Transform via Filter Banks

From the previous section, we see that the dyadic wavelet transform can be used as a way to carry out multiresolution analysis. But such a multiresolution decomposition, or mathematically, a projection onto the different subspaces, needs to be carried out efficiently. A fast method based on filter banks does exactly this.

From the nesting of subspaces (equations (3.3.4-2) and (3.3.4-4)), a fundamental relationship between the scaling functions of two consecutive scales exists:

$$s(t) = \sum_{k=-\infty}^{\infty} g(k)s(2t-k), \qquad (3.3.5\text{-}1)$$

which is known as a *two-scale difference equation*. From (3.3.4-5), we deduce that $W_j \subset V_{j+1}$, or

$$\psi(t) = \sum_{k=-\infty}^{\infty} h(k)s(2t-k). \qquad (3.3.5\text{-}2)$$

Burrus and Gopinath [22] have shown how a filter bank interpretation of the dyadic wavelet transform results from a manipulation of equations (3.3.5-1) and (3.3.5-2). (Hereafter, for the purpose of clarity, the taking of the dyadic wavelet transform implies multiresolution decomposition and vice versa.)

Now let us see how a signal decomposition can be carried out with filter banks. Since $W_j \oplus V_j = V_{j+1}$, if

$$f(t) \in V_{j+1}$$

then

$$f(t) = \sum_{k=-\infty}^{\infty} a_{j+1} 2^{j+1/2} s(2^{j+1}t-k), \qquad (3.3.5\text{-}3)$$

or

$$f(t) = \sum_{k=-\infty}^{\infty} a_j(k)2^{j/2} s(2^j t-k) + \sum_{k=-\infty}^{\infty} b_j(k)2^{j/2} \psi(2^j t-k). \qquad (3.3.5\text{-}4)$$

Substituting $t = 2^j t - k$ into equations (3.3.5-1) and (3.3.5-2), we have

$$s(2^j t-k) \;=\; \sum_{n=-\infty}^{\infty} g(n-2k)\,s(2^{j+1}t-n) \qquad (3.3.5\text{-}5)$$

and

$$\psi(2^j t-k) \;=\; \sum_{n=-\infty}^{\infty} h(n-2k)\,s(2^{j+1}t-n). \qquad (3.3.5\text{-}6)$$

To solve for the coefficients $a_j(k)$ in (3.3.5-4), we need to multiply both sides of (3.3.5-4) by $2^{j/2}\,s(2^j t-k)$ and integrate with respect to t. If the bases $s_{j,k}$ and $\psi_{j,k}$ are orthonormal, then because they are also orthogonal complements of each other, we obtain

$$\int_{-\infty}^{\infty} f(t)\,2^{j/2}\,s(2^j t-k)\,dt \;=\; a_j(k). \qquad (3.3.5\text{-}7)$$

If $s(2^j t-k)$ is replaced with (3.3.5-5), the coefficients can be generated recursively as

$$a_j(k) \;=\; \sum_{n=-\infty}^{\infty} g(n-2k)\,a_{j+1}(n). \qquad (3.3.5\text{-}8)$$

The coefficients $b_j(k)$ can be solved for in a similar manner. After multiplying (3.3.5-4) by $2^{j/2}\,\psi(2^j t-k)$ and integrating with respect to t, we have

$$\int_{-\infty}^{\infty} f(t)\,2^{j/2}\,\psi(2^j t-k)\,dt \;=\; b_j(k). \qquad (3.3.5\text{-}9)$$

Further manipulation leads to

$$b_j(k) \;=\; \sum_{n=-\infty}^{\infty} h(n-2k)\,a_{j+1}(n). \qquad (3.3.5\text{-}10)$$

From the above derivation, we see that equations (3.3.5-8) and (3.3.5-10) provide us with a fast way to decompose a signal into a series of lower resolution components. In fact, for a signal represented at resolution $j+1$, the lower resolution components can be obtained by filtering and downsampling by a factor of two. This process can be repeated to obtain lower and lower resolution components (Fig. 3.3.5-1).

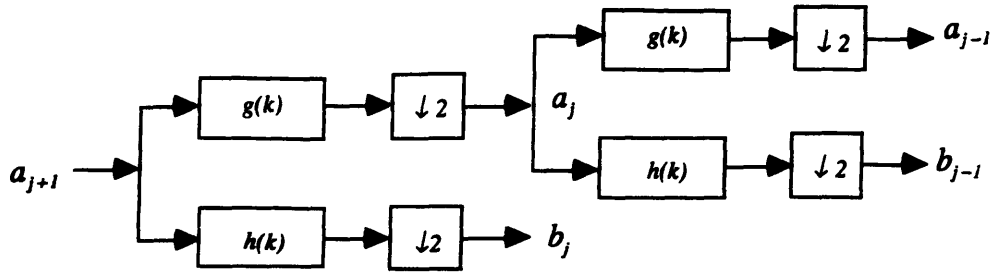


Figure 3.3.5-1: Generation of lower resolution components with filter banks.

An equally attractive way to reconstruct a signal from its lower resolution components also exists. Setting equation (3.3.5-3) equal to (3.3.5-4) yields

$$\sum_{k=-\infty}^{\infty} a_{j+1}(k)2^{j+1/2}\, s(2^{j+1}t-k) = \sum_{k=-\infty}^{\infty} a_j(k)2^{j/2}\, s(2^j t-k) + \sum_{k=-\infty}^{\infty} b_j(k)2^{j/2}\, \psi(2^j t-k). \qquad (3.3.5\text{-}11)$$

If we substitute equations (3.3.5-1) and (3.3.5-2) into the right side of (3.3.5-11), multiply both sides of (3.3.5-11) by $s(2^{j+1}t-k)$ before integrating with respect to t, we have

$$a_{j+1}(k) = \sum_{n=-\infty}^{\infty} a_j(n)g(k-2n) + \sum_{n=-\infty}^{\infty} b_j(n)h(k-2n). \qquad (3.3.5\text{-}12)$$

Hence, a signal represented at resolution $j+1$ can be obtained by first upsampling by a factor of two the coefficients at resolution j and then filtering with the same filters that are used during the decomposition stage. The procedure can be repeated to rebuild the signal at higher and higher resolutions (Fig. 3.3.5-2).

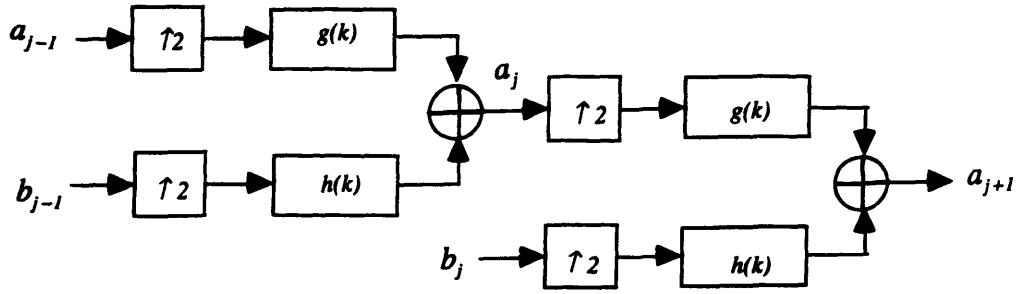


Figure 3.3.5-2 : Generation of higher resolution components with filter banks.

This fast multiresolution decomposition-reconstruction algorithm is attributed to Mallat, and a more detailed derivation can be found in [22], [24], and [36].

To further study how the filter implications of equations (3.3.5-1) and (3.3.5-2) are related to subband coding techniques, let us first take the Fourier transform of both sides of (3.3.5-1) such that

$$\hat{s}(\omega) = \frac{1}{2}\hat{g}(e^{i\frac{\omega}{2}})\hat{s}(\frac{\omega}{2}), \qquad (3.3.5\text{-}13)$$

where

$$\hat{g}(e^{i\omega}) = \sum_{k=-\infty}^{\infty} g(k)e^{-i\omega k}.$$

If the family of functions, $s(t-k)$, $k \in Z$, is orthogonal with respect to each other, then according to the Poisson formula [18], equation (3.3.5-13) implies that the following holds in the frequency domain:

$$\sum_{k=-\infty}^{\infty}|\hat{s}(\omega + 2k\pi)|^2 = 1, \qquad (3.3.5\text{-}14)$$

which also means that

$$\sum_{k=-\infty}^{\infty}|\hat{s}(2\omega + 2k\pi)|^2 = 1. \qquad (3.3.5\text{-}15)$$

Let us now examine the implications of equations (3.3.5-14) and (3.3.5-15). First, replace ω in (3.3.5-13) with 2ω so that

$$\hat{s}(2\omega) = \frac{1}{2}\hat{g}(e^{i\omega})\hat{s}(\omega). \tag{3.3.5-16}$$

Next, replace ω in (3.3.5-16) with $\omega + k\pi$ to obtain

$$\hat{s}(2\omega + 2k\pi) = \frac{1}{2}\hat{g}(e^{i(\omega+k\pi)})\hat{s}(\omega + k\pi). \tag{3.3.5-17}$$

After taking the absolute value, squaring, and summing over all k on both sides of (3.3.5-17), we have

$$\sum_{k=-\infty}^{\infty}|\hat{s}(2\omega + 2k\pi)|^2 = \frac{1}{4}\sum_{k=-\infty}^{\infty}|\hat{g}(e^{i(\omega+k\pi)})|^2\,|\hat{s}(\omega + k\pi)|^2. \tag{3.3.5-18}$$

Substituting (3.3.5-15) into the left side of (3.3.5-18) yields

$$4 = \sum_{k=-\infty}^{\infty}|\hat{g}(e^{i(\omega+k\pi)})|^2\,|\hat{s}(\omega + k\pi)|^2. \tag{3.3.5-19}$$

Since $\hat{g}(e^{i\omega})$ is a 2π-periodic function, we can rewrite (3.3.5-19) in terms of even and odd k's, so that

$$\begin{aligned} 4 = &\sum_{k=-\infty}^{\infty}|\hat{g}(e^{i(\omega+2k\pi)})|^2\,|\hat{s}(\omega + 2k\pi)|^2 + \\ &\sum_{k=-\infty}^{\infty}|\hat{g}(e^{i(\omega+(2k+1)\pi)})|^2\,|\hat{s}(\omega + (2k+1)\pi|^2. \end{aligned} \tag{3.3.5-20}$$

When (3.3.5-14) and (3.3.5-15) are substituted into (3.3.5-20), we have

$$4 = |\hat{g}(e^{i\omega})|^2 + |\hat{g}(e^{i(\omega+\pi)})|^2, \tag{3.3.5-21}$$

which is the defining equation for a class of perfect reconstruction quadrature mirror filters (QMF), which have been extensively studied and applied in subband coding [30]. Since QMF's deal with one low-pass and one band-pass filter, we need to identify these filters.

Multiplication of both sides of (3.3.5-1) by dt followed by an integration with respect to t gives

$$2 = \sum_{k=-\infty}^{\infty} g(k), \tag{3.3.5-22}$$

or

$$\hat{g}(e^{i\omega})\big|_{\omega=0} = 2, \tag{3.3.5-23}$$

which says that $\hat{g}(e^{i\omega})$ is actually the low-pass filter. A similar analysis shows that $\hat{h}(e^{i\omega})$ is the band-pass filter. From the QMF relationship, the filter sequences, $g(k)$ and $h(k)$, are related by

$$h(k) = (-1)^k g(-k+1). \qquad (3.3.5\text{-}24)$$

So far we have seen that decomposing and reconstructing a signal involve simple filtering operations. An interesting filtering interpretation of the scaling and wavelet functions also exists. From the orthogonality of the scaling equation with its translates along the time axis, Mallat in [27] has shown that the scaling function is actually a low-pass filter. Likewise, due to the orthogonality of the mother wavelet with its translates, he has found that the wavelet function also represents a band-pass filter. Thus, because $s_{j,k}(t)$ spans V_j and $\psi_{j,k}(t)$ spans W_j, whenever we talk about mapping a function onto the V_j and W_j subspaces, this simply corresponds to low-pass and band-pass filtering it, respectively, in signal processing terminology.

3.3.6 Dyadic Wavelet Basis Construction

One of the simplest ways to construct a dyadic wavelet basis is to start with the coefficients, $g(k)$, of the two-scale difference equation, (3.3.5-1). Having the coefficients, one needs to find a scaling function that satisfies (3.3.5-1). The overall heuristic approach is very similar to solving a differential equation.

Burrus and Gopinath [22] have developed a much more structured method. By using the fact that the scaling function and its integer translates are orthogonal, they imposed another constraint on the coefficients. They first select the number of non-zero values in $g(k)$. From the number of coefficients supported by $g(k)$ and the orthogonality constraint, they solve a system of equations based on equation (3.3.5-1) for $g(k)$.

Once $g(k)$ and $s(t)$ are determined, $h(k)$ and the wavelet function, $\psi(t)$, can be obtained by equations (3.3.5-24) and (3.3.5-2), respectively. To see if $\psi(t)$ is a mother wavelet, one needs to check whether $\int_{-\infty}^{\infty} \psi(t)dt = 0$ holds.

Let us now look at some popular wavelets. The simplest one is the Haar wavelet. The Haar scaling function is

$$s_{Haar}(t) = \begin{cases} 1, & 0 \le t < 1 \\ 0, & \text{otherwise} \end{cases} \qquad (3.3.6\text{-}1)$$

which is also known as a box function (Fig. 3.3.6-1).

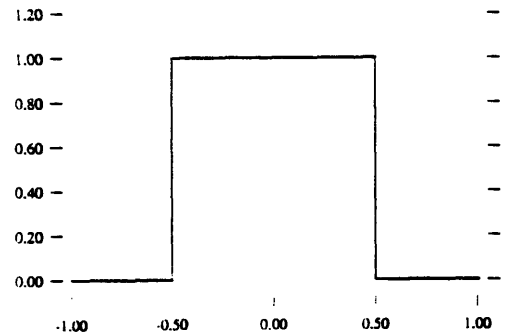


Figure 3.3.6-1: Box function, scaling function for the Haar wavelet.

In this case,

$$g(0) = 1,$$
$$g(1) = 1,$$

so that

$$h(0) = 1,$$
$$h(1) = -1,$$

and the Haar wavelet results (Fig 3.3.6-2).

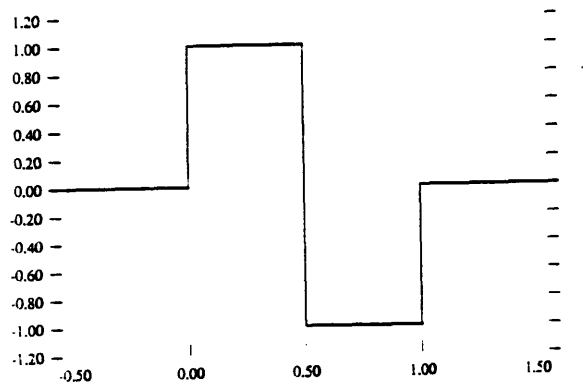


Figure 3.3.6-2: Haar wavelet.

From Fig. 3.3.6-2, one can easily see that the Haar wavelets are orthogonal across both shifts and scales so that they form an orthonormal basis for $L^2(\Re)$. In addition, the Haar wavelet basis has compact support, which makes it a potentially useful tool for signal analysis.

Daubechies [23] has discovered another class of compactly supported orthonormal wavelets that span $L^2(\Re)$ - Strang [28] calls them "an inspiration of Daubechies." The scaling function is shown in Fig. 3.3.6-3, and

$$g(0) = \frac{1+\sqrt{3}}{4},$$
$$g(1) = \frac{3+\sqrt{3}}{4},$$
$$g(2) = \frac{3-\sqrt{3}}{4},$$
$$g(3) = \frac{1-\sqrt{3}}{4}.$$

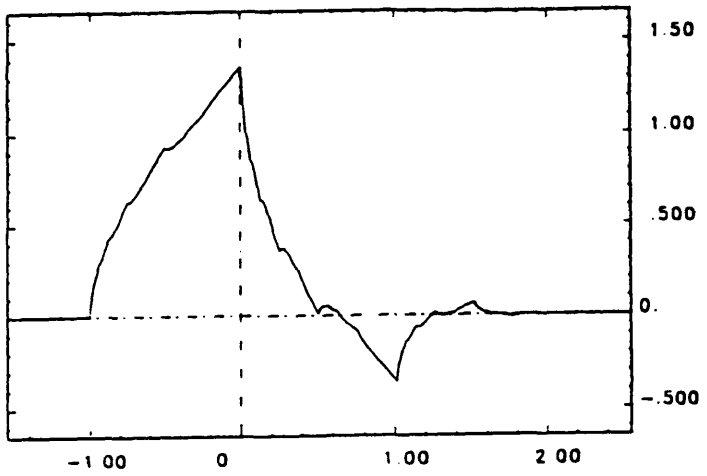


Figure 3.3.6-3: Daubechies scaling function. (After Daubechies [23])

Fig. 3.3.6-4 shows the resulting mother wavelet.

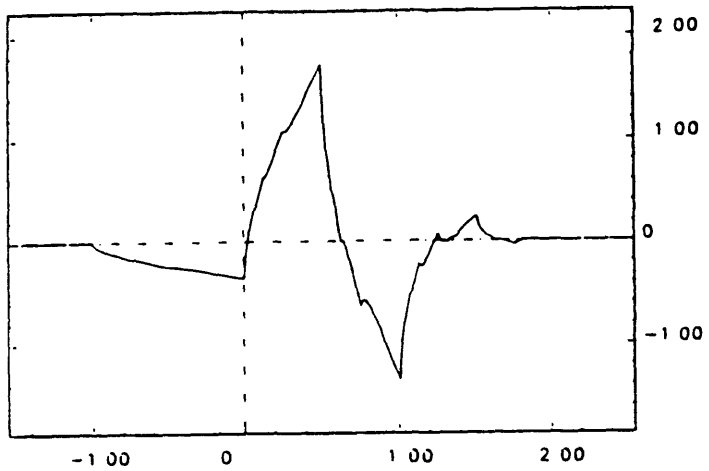


Figure 3.3.6-4: Daubechies wavelet function.(After Daubechies [23])

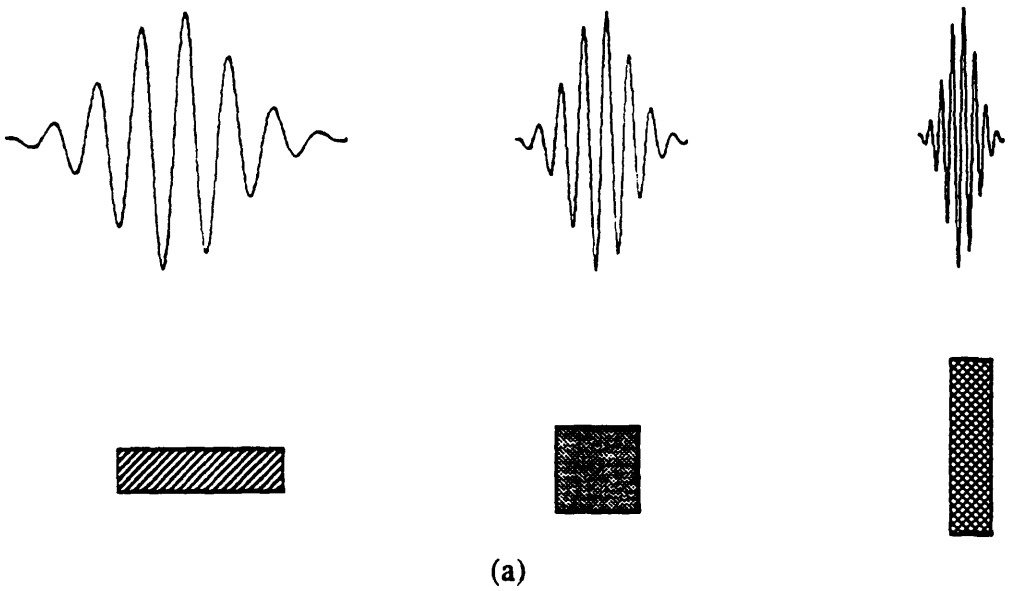
While both the Haar wavelet and the Daubechies wavelet have compact support and are orthonormal across both translates and scales, they are, however, not smooth. In signal analysis applications, particularly in speech analysis in which the signal is relatively smooth, this lack of smoothness, or irregularity, is highly inappropriate. This unfavorable property is the price paid for the favorable properties of orthogonality and compact support [28]. In the analysis of signals which we have no a priori knowledge about, compact

support is absolutely essential. Thus, a trade-off must be made between orthogonality and smoothness, while compact support must be retained.

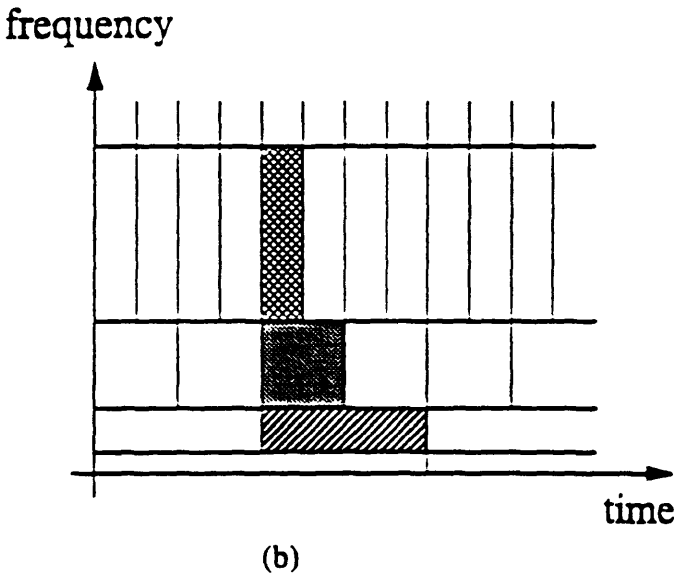
3.4 Wavelet Analysis vs. Fourier Analysis

Now that we have seen the basics of wavelet analysis and Fourier analysis, let us examine their similarities and differences.

- Basis function: The similarities and differences among the definitions of the FT (equation (2.3-1)), STFT (equation (2.4-1), and CWT (equation (3.3.1-2)) are very obvious. Unlike that of the FT and STFT, the analyzing function of the CWT is arbitrary.
- Signal expansion: For the STFT, the cost of obeying the uncertainty principle is that both the duration and the bandwidth of the basis function are constant (Fig. 2.4.2-3a). However, from the definition of the CWT, we see that such an inflexibility can be avoided, while the uncertainty principle is still obeyed. To see this more clearly, let us examine the basis functions of the dyadic wavelet transform, $\psi_{j,k}(t) = 2^{j/2}\,\psi(2^j(t - 2^{-j}k))$. For j large and positive, the basis function's time duration is small so that the time shifts are also small (Fig. 3.4-1a). Such a feature will be extremely helpful in the analysis of high frequency signals. On the other hand, for j large but negative, the basis function's time duration is large, and the time shifts are also large (Fig. 3.4-1a). This allows a reliable detection of truly low frequency components. By sweeping the scale and shift parameters through a set of values, we can accurately determine the existence of fast-changing, slow-changing, or both kinds of components embedded in a signal. We can also determine the time at which these components exist. In conclusion, flexibility is what separates the wavelet techniques from the Fourier techniques.



(a)



(b)

Figure 3.4-1: Wavelet transform basis functions and time-frequency resolution cells. (After Vetterli and Herley [36]).
(a) Basis Functions.
(b) Time-frequency resolution cells.

- Multiresolution capabilities: Since the wavelet itself can be interpreted as a bandpass filter, the entire dyadic wavelet transform (equation (3.3.3-1)) can be seen as an octave band filtering operation. In the case of the STFT, the time-frequency plane is covered by rigid constant band time-frequency cells. Let us now look at how the plane is covered differently in the case of the dyadic wavelet transform.

  From the notion of scale, we see that the frequency domain is covered octave by octave. And since the scale parameter j is also included in the time-shift width, $2^{-j}k$, we see that for great frequency resolution, inferior time resolution results (i.e., the time width is large). On the other hand, for great time resolution, inferior frequency resolution results (i.e., the frequency bandwidth is large). The key here is that we can *control* the resolution. If our original choice of a certain kind of resolution turns out to be unsatisfactory, adaptive or feedback schemes can be installed to obtain improvements.

  Fig. (3.4-1b) shows the dynamic time-frequency resolution capabilities of the dyadic wavelet transform. The main difference between Fig. (2.4.1-3b) and Fig. (3.4-1b) is that a fundamentally different trade-off is made while the uncertainty principle is still obeyed. In the wavelet transform, the time-duration frequency-bandwidth product is still constant, but the time duration, Δt, and the frequency bandwidth, $\Delta\omega$, are allowed to change.

3.5 Why the Dyadic Wavelet Transform?

A reason for the tremendous interest in the wavelet transform is that it is a very good model of the main signal processing capabilities of the human visual and auditory systems, which have evolved into optimal image and speech processors, respectively. If

one considers the proficiency with which these systems process signals, one may begin to see why such great interest is warranted.

To start things off, let us look at a model of the signal processing mechanism of the human visual system. Interested in applying the wavelet methods in computer vision, Mallat [33] has thoroughly examined the experiments elucidating the physiology and psychophysics of vision. Experimental results have shown that the human visual system processes images by carrying out a multifrequency channel decomposition: in other words, images are broken down by separate frequency channels before they are analyzed. In addition, the frequency bandwidth of each of these channels has been found to be constant on a logarithmic scale.

As for the auditory system, a very similar kind of analysis occurs. The main hearing mechanism can be modeled as a bank of band-pass filters which perform a log-linear analysis of sounds [67]. Such an octave-band filtering model has already been employed in high fidelity broadband audio coding to obtain transparent sound quality.

The models for both of the above systems consist of a bank of filters that have constant bandwidth on a logarithmic scale. Examining the time-frequency resolution of the wavelet transform (Fig. (3.4-1b)), we see that the entire vertical frequency axis is also covered by a bank of octave-band filters (Fig. 3.5 -1).

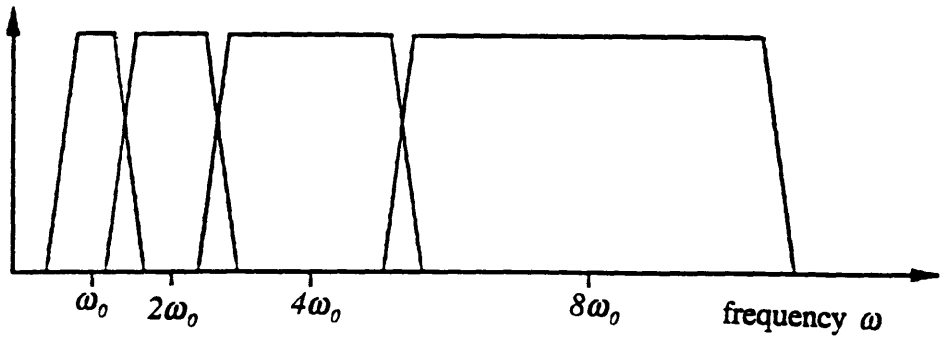


Figure 3.5-1: Octave-band filter banks of the dyadic wavelet transform. (After Rioul and Vetterli [6]).

In contrast to the STFT, whose filter bandwidths are constant on a linear scale (Fig. 2.4.1-3b), the dyadic wavelet transform consists of filter bandwidths which are constant on a logarithmic scale, or constant-Q filters, such that

$$\frac{\Delta\omega}{\omega} = constant.$$

Now that we have seen how the dyadic wavelet transform adequately models the main signal processing mechanisms in the human visual and auditory systems, we can see that it has great potential for success in many signal processing applications, especially in image processing and speech processing. However, the degree of success is inherently dependent on the wavelets themselves. The next chapter introduces a class of smooth but non-orthogonal wavelets built from splines. Chapter 5 deals with the use of the wavelet transform based on B-spline wavelets in the particular application of pitch detection.

Chapter 4 Basic Spline Wavelet Transform

4.1 Introduction

This chapter discusses the wavelet transform based on wavelets constructed from basic spline functions. Section 4.2 provides a historical background of these functions and their applications in signal processing. Section 4.3 covers the spline functions and their properties that will be useful in the formulation of our transform. In Section 4.4, we derive the transform and discuss its usage.

4.2 Historical Background

The basic spline wavelet transform culminates from the areas of spline interpolation and approximation in applied mathematics and wavelet analysis in signal processing. Before examining how the transform has come about, let us familiarize ourselves with the terminology.

The term *spline* was coined by Schoenberg in 1946 to denote a function that is defined piecewise: each of these pieces was "tied" to its neighbors at the *knot points*. The designation was justified, for, in fact,

> "A spline is a mechanical instrument consisting
> of a flexible rod which can be used to draw
> smooth curves through prescribed points." - [71].

If each of these pieces is represented by a polynomial, the function is known as a *polynomial spline*, which has played a very important role in interpolation and approximation theory and numerical analysis. An early application of these splines was in ballistics in which a smooth function was constructed from a given set of data points.

Because polynomial splines have a relatively simple form and are therefore easy to manipulate, they have been applied in many areas of signal processing. An important application is in image interpolation in which an efficient resolution adaptive scheme is often needed on modern workstations so that an user can view an image more interactively [49]. Over the years applications of polynomial splines have been extensively studied:

- Horowitz [48] (1974) studied the power spectral effects of spline interpolators and found they were much more favorable than those of conventional zero-order holds.
- Liou [47] (1976) devised a curve-fitting algorithm based on third-order, or cubic, splines with equally spaced data points and given end conditions.
- Hou and Andrews [45] (1978) applied a particular type of spline, the basic spline (or B-spline), as an image interpolator. They found that the spline-based interpolation schemes were better than other interpolation methods for image magnification and reduction and noise smoothing.
- Chen and deFigueiredo [75] (1985) proposed generalized spline interpolation schemes based on partial differential equation (PDE) models of an image signal.
- Sankar and Ferrari [76] (1988) proposed an efficient B-spline interpolation scheme using running average filters and pipelined architectures.
- Unser *et al.* ([49], [50], [51], [52], [53], [54], [55]) studied the digital signal processing capabilities of the B-spline functions.

Though polynomial splines themselves have a long and rich history, the idea of using splines as a wavelet basis was not proposed until recently. However, there were some mathematical methods that did employ splines as basis functions. For example, in the Haar transform, the basis functions were actually zero-order splines.

Recently Battle and Lemarie independently introduced a class of orthogonal spline wavelets [77]. But by being orthogonal, the basis functions have exponential decay, or non-compact support, and became unattractive for non-stationary signal analysis. Building on Mallat's work on multiresolution signal decomposition and wavelet analysis, Chui and Wang ([58], [59]) and Unser *et al.* [57] have independently proposed the use of compactly supported, non-orthogonal B-splines as wavelets.

4.3 Basic Spline Functions and Polynomial Spline Interpolation

Before we analyze the special case of the B-spline, let us first study the polynomial splines, which are polynomials of degree n in each of the intervals $[t_k, t_{k+1})_{k \in Z}$ that are smoothly connected to one another at the knot points so that the entire function is in C^{n-1} (i.e., the space of continuous functions with continuous derivatives up to order n - *1*).

The interest in the B-spline comes from Schoenberg's discovery that any order n polynomial spline, $s^n(t)$, can be represented by a linear combination of shifted nth-order B-splines, $\beta^n(t)$ [56]:

$$s^n(t) = \sum_{k=-\infty}^{\infty} c(k)\,\beta^n(t-k), \qquad (4.3\text{-}1)$$

where $t \in \Re$ and $k \in Z$. In other words, any polynomial spline can be uniquely characterized by its B-spline coefficients, $c(k)$. (Here, we are dealing specifically with polynomial splines with integral knot points, in which case the orders, n, are restricted to be *odd* [56].) The B-splines themselves are constructed by repeated convolution of a central zero-order B-spline, $\beta^0(t)$, which is a characteristic function defined on the interval $[\frac{-1}{2}, \frac{1}{2})$ (Fig. 4.3-1a),

$$\beta^0(t) = \begin{cases} 1, & t \in [-0.5, 0.5) \\ 0, & \text{otherwise} \end{cases} \qquad (4.3\text{-}2)$$

A first-order B-spline (Fig. 4.3-1b) is generated by a convolution of two zero-order B-splines,

$$\beta^1(t) = \beta^0(t) * \beta^0(t), \tag{4.3-3}$$

and a second-order B-spline (Fig. 4.3-1c) is constructed with another convolution. In general, a *n*th-order B-spline can be generated by

$$\begin{aligned} \beta^n(t) &= \beta^{n-1}(t) * \beta^0(t) \\ &= \underbrace{\beta^0(t) * \beta^0(t) * \ldots * \beta^0(t)}_{n+1 \text{ times}} \end{aligned} \tag{4.3-4}$$

In terms of a symmetrical B-spline generating kernel, Schoenberg formulates an explicit expression for a *n*th-order B-spline as [53]

$$\beta^n(t) = \sum_{r=0}^{n+1} \frac{(-1)^n}{n!} \binom{n+1}{r} \left(t + \frac{(n+1)}{2} - r\right)^n u\left(t + \frac{(n+1)}{2} - r\right), \tag{4.3-5}$$

where $u(t)$ is the unit step function and $\binom{n+1}{r}$ stands for the binomial coefficients:

$$\binom{n+1}{r} = \frac{(n+1)!}{(n+1-r)!\ r!}$$

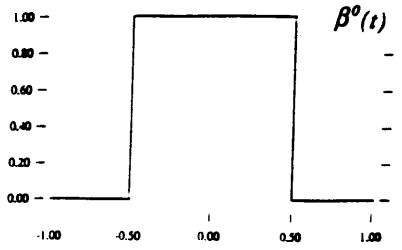
The attractiveness of the B-splines is due to the following properties:

- Compact support: Among all *n*th-order polynomial splines, the B-spline has minimal support [72].
- Smooth: An examination of Figs. 4.3-1a, b, c shows that the degree of smoothness increases with the number of convolutions. (Here, the degree of smoothness of a function is determined by its number of continuous derivatives. Hence, a third-order B-spline is smoother than a second-order B-spline because the former has one more continuous derivative.)
- Simple form: Equation (4.3-5) furnishes the individual values of a B-spline of any order. The equivalent of (4.3-4) in the frequency domain is

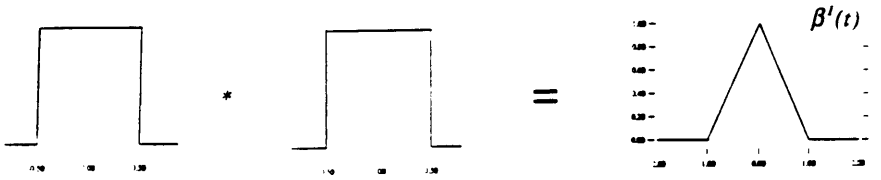
  $$\hat{\beta}^n(f) = [sinc(f)]^{n+1}, \tag{4.3-6}$$

  where

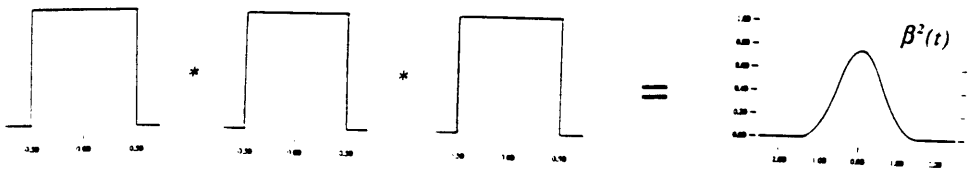
  $$sinc(x) = \frac{\sin(\pi x)}{\pi x} \quad .$$



(a)



(b)



(c)

Figure 4.3-1: B-spline construction.

(a) Central zero-order B-spline, $\beta^0(t)$.

(b) First-order B-spline, $\beta^1(t)$.

(c) Second-order B-spline, $\beta^2(t)$.

At this point only continuous B-spline functions have been considered. The discrete counterpart, obtained by sampling the continuous version, is

$$b^n(k) = \beta^n(t)\Big|_{t=k}, \tag{4.3-7}$$

where $k \in Z$, and a discrete nth-order B-spline expanded by an integral factor of m is

$$b_m^n(k) = \beta^n(\frac{k}{m}). \tag{4.3-8}$$

An important point needs to be considered here. From equation (4.3-1), we see that any polynomial spline can be represented by B-splines so that any function that can be approximated with polynomial splines can also be expanded in terms of B-splines. Therefore, for a nth-order polynomial spline, if there exists a sequence $x(k)$ such that a function $f(t)$ can be represented as

$$f(t) = \sum_{k=-\infty}^{\infty} x(k)\, s^n(t-k), \tag{4.3-9}$$

then there must exist a sequence $y(k)$ such that

$$f(t) = \sum_{k=-\infty}^{\infty} y(k)\, \beta^n(t-k). \tag{4.3-10}$$

4.4 Basic Spline Wavelet Transform

The basic spline wavelet transform, developed by Unser *et al.* [57], is an efficient way to evaluate the continuous wavelet transform by approximating both the input signal and the analyzing wavelet with a sum of shifted B-splines. The CWT is of particular interest here because, unlike the DWT, it is both time-shift invariant and scale-change invariant.

Since we are working with analyzing wavelets which are real functions, complex conjugation in the definition of the CWT (equation (3.3.1-1)) can be ignored so that

$$|a|^{1/2}\,(CWT\, f)(b,a) = \int_{-\infty}^{\infty} f(t)\, \psi(\frac{t-b}{a})dt. \tag{4.4-1}$$

In the DWT, the scale and shift parameters are chosen as $a = M^j$ and $b = kM^jN$, where $M, N \in \Re$ and $j, k \in Z$. For the special case of the dyadic wavelet transform, the parameters become $a = 2^{-j}$ and $b = k2^{-j}$, in which case critical sampling results and Mallat's fast QMF-based filtering algorithm can be used to evaluate the transform. Taking advantage of the strong approximation capabilities of B-splines, let us now derive a transform that is much less computationally burdensome than that defined in (4.4-1).

In most applications, it is highly impractical to used a scale parameter that is real. So let us restrict the scale parameter to be integral, i.e., $a = m$, where $m \in Z$. Using this constraint and recognizing something very similar to convolution in (4.4-1), we introduce a revised but related transform, CWT_2, based on filtering:

$$(CWT_2\, f)(b,m) \;=\; \psi(\frac{b}{m}) * f(b), \tag{4.4-2}$$

where $m \in Z$ and $b \in \Re$. Further simplification can be achieved if the two functions on the right-hand side of (4.4-2) are approximated with B-splines.

Now assume that the continuous function, $f(t)$, can be approximated by n_1th-order B-splines so that

$$f(t) \;=\; \sum_{k=-\infty}^{\infty} p(k)\, \beta^{n_1}(t-k), \tag{4.4-3}$$

where $t \in \Re$, $k \in Z$, and $\beta^0(t)$ is the central B-spline defined in (4.3-2). For a discrete function, $f(k)$, the expansion becomes

$$\begin{aligned} f(k) \;&=\; \sum_{y=-\infty}^{\infty} p(y)\, b^{n_1}(k-y) \\ &=\; p(k) * b^{n_1}(k), \end{aligned} \tag{4.4-4}$$

where $b^{n_1}(k)$ is a discrete B-spline defined in (4.3-7).

To find a B-spline approximation for $\psi(\frac{t}{m})$, we start with an approximation for $\psi(t)$ with B-splines of order n_2:

$$\psi(t) \;=\; \sum_{k=-\infty}^{\infty} q(k)\, \beta^{n_2}(t-k). \tag{4.4-5}$$

In the case of $\psi(\frac{t}{m})$, the expansion simply becomes

$$\psi(\frac{t}{m}) = \sum_{k=-\infty}^{\infty} q(k)\,\beta^{n_2}(\frac{t}{m}-k). \qquad (4.4\text{-}6)$$

To simplify (4.4-6), we make use of the fact that an *n*th-order B-spline (*n odd*) expanded by an integral factor of *m* is still a B-spline with integral knot points ([56], [57]). Mathematically, this translates into

$$\beta^n(\frac{t}{m}) = \sum_{y=-\infty}^{\infty} u_m^n(y)\,\beta^n(t-y), \qquad (4.4\text{-}7)$$

where, provided that *n* and *m* are not both even (always the case since we are working with *odd n*), the Z-transform of $u_m^n(k)$ is

$$U_m^n(z) = \frac{z^{k_0}}{m^n}(\sum_{k=0}^{m-1} z^{-k})^{n+1}, \qquad (4.4\text{-}8)$$

where $k_0 = \frac{(n+1)(m-1)}{2}$. A detailed proof of (4.4-7) can be found in [57].

Now let us simplify equation (4.4-6). Substituting (4.4-7) into (4.4-6) gives

$$\psi(\frac{t}{m}) = \sum_{k=-\infty}^{\infty} q(k) \sum_{y=-\infty}^{\infty} u_m^{n_2}(y)\,\beta^{n_2}(t-mk-y). \qquad (4.4\text{-}9)$$

Using $r = mk + y$, we can rewrite (4.4-9) as

$$\psi(\frac{t}{m}) = \sum_{k=-\infty}^{\infty} q(k) \sum_{r=-\infty}^{\infty} u_m^{n_2}(r-mk)\,\beta^{n_2}(t-r), \qquad (4.4\text{-}10)$$

which, after the order of summations is switched, yields

$$\begin{aligned}\psi(\frac{t}{m}) &= \sum_{r=-\infty}^{\infty} (\sum_{k=-\infty}^{\infty} q(k)\,u_m^{n_2}(r-mk))\,\beta^{n_2}(t-r) \\ &= \sum_{r=-\infty}^{\infty} ([q(r)]_{\uparrow m} * u_m^{n_2}(r))\,\beta^{n_2}(t-r),\end{aligned} \qquad (4.4\text{-}11)$$

where $[\bullet]_{\uparrow m}$ designates upsampling by an integral factor of *m*. From (4.3-4), it can be deduced that

$$\beta^{n_1}(t) * \beta^{n_2}(t) = \beta^{n_1+n_2+1}(t). \qquad (4.4\text{-}12)$$

Now if we substitute equations (4.4-3) and (4.4-11) into (4.4-2) and then use relation (4.4-12) to simplify, we obtain a transform that can be expressed in terms of $(n_1 + n_2 + 1)$th-order B-splines:

$$(CWT_2 f)(b,m) = \sum_{k=-\infty}^{\infty} ([q]_{\uparrow m} * u_m^{n_2} * p)(k)\,\beta^{n_1+n_2+1}(b-k). \qquad (4.4\text{-}13)$$

In addition, if only integral shifts are considered (i.e., $b \in Z$), the transform can be simply expressed in terms of convolutions:

$$(CWT_2 f)(b,m)\big|_{b=k} = [q]_{\uparrow m} * u_m^{n_2} * p * b^{n_1+n_2+1}(k). \qquad (4.4\text{-}14)$$

If the input signal is approximated with linear splines (i.e., $n_1 = 1$), then $p(k) = f(k)$, and the final form of the wavelet transform based on B-splines, or basic spline wavelet transform (BSWT), results:

$$(BSWT\ f)(k,m) = [q]_{\uparrow m} * u_m^{n_2} * b^{n_2+2} * f(k). \qquad (4.4\text{-}15)$$

Besides allowing us a convenient way to view a signal at different integral scales and discrete-time shifts, the transform defined in equation (4.4-15) lets us design a wavelet for a specific application. The sequence $q(k)$ permits us to choose a wavelet's shape and support, while n_2 lets us control the degree of smoothness. These issues and their importance will be explored in detail in the next chapter in which the transform is used to detect pitch.

Chapter 5 Pitch Detection with the Basic Spline Wavelet Transform

5.1 Introduction

In this chapter, we investigate the important factors that must be considered in pitch detection with basic spline wavelets. The performance of these wavelets in adverse environments is evaluated. Section 5.2 covers the basic issues involved in the pitch detection problem. In Section 5.3, we discuss the experiments based on the designed spline wavelets. Section 5.4 presents and discusses the performance of the experimental wavelets for various kinds of test data.

5.2 The Pitch Detection Problem

Reliable pitch detection of a speech signal has been an elusive problem in speech research for many decades. The problem is especially formidable in a noisy environment. To see why this is so, we first need to look into the human speech production mechanism and see how pitch is produced.

Human speech sounds are produced when inhaled air is expelled from the lungs into the vocal tract and then radiated as an acoustic wave from the mouth, nostrils, or both. Fig. 5.2-1 shows a diagram of the main anatomical parts of the vocal system. Sounds are classified as voiced, unvoiced, or mixed. What separates one from the others is the presence, or lack of it, of a certain repetitive structure in the signal. In voiced sounds, for

example, waveform periodicity is present; for unvoiced, it is not. Let us see where this periodicity comes from.

A slit-like orifice between the vocal cords [88], the glottis (Fig. 5.2-2) is the site of pitch generation. Its function can be modeled as a valve. As air is forced from the lungs, pressure builds behind the closed glottis, and eventually, overcomes resistance of the vocal cords. The glottis opens and air flows into the vocal tract. As the pressure dissipates, the vocal cords' muscle tension causes the glottis to quickly close. When the next gush of air is expelled from the lungs and pressure is regenerated, the glottal opening and closing processes repeat. These processes translate into waveform periodicity in the form of repetitive pitch pulses ([15], [80]).

To see what voicing means in a speech signal, let us examine Fig. 5.2-3a, which shows a time plot of the word "appetite" sampled at 8 kHz, and Fig. 5.2-3b, which shows 320 samples (40 ms) of the voiced sound \a\ at the front of the word. Periodicity is very evident in the latter figure. The problem of pitch detection is the determination of this periodicity, or pitch period, defined as the time between two consecutive glottal closures, which show up as large amplitude spikes in a signal. (To clarify the terminology used here, we use pitch period as a representation in time, while pitch alone is a representation in either time or frequency.)

Interest in accurate pitch detection comes from the important role the pitch period plays in many applications involving speech perception and intelligibility :

- In analysis-synthesis low bit-rate speech coding, accurate pitch determination is absolutely essential for high-quality signal reproduction because the human auditory system is very perceptive to pitch and its variations [88].
- Atal [70] (1968) found that pitch could be an effective speaker-discriminating parameter in automatic speaker recognition.
- In many cases, pathological phenomena in the speech production system can be diagnosed by studying a corrupted spoken sound's pitch period.

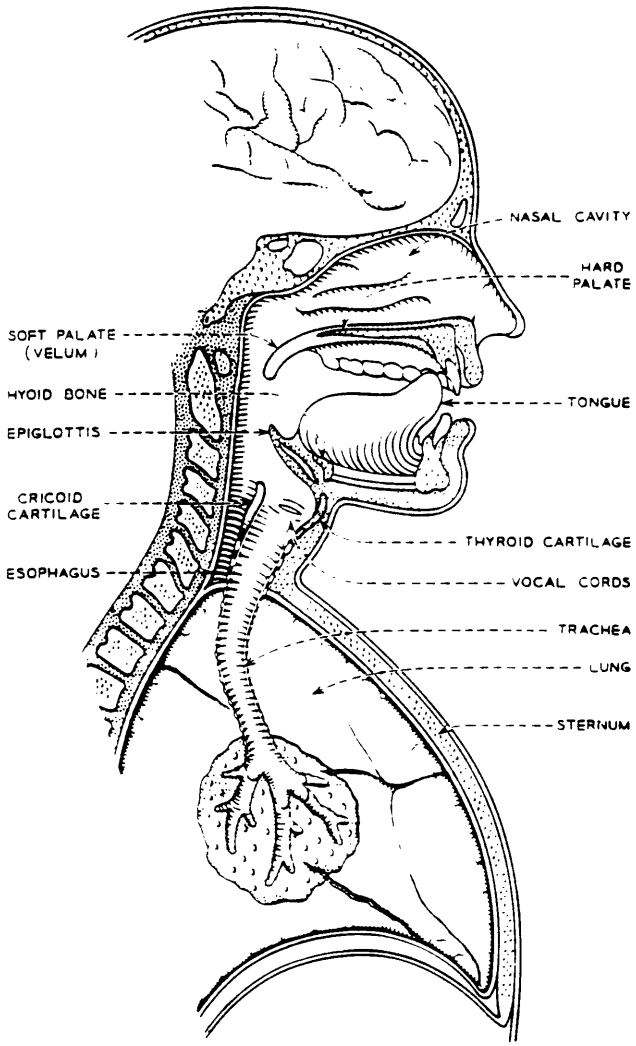
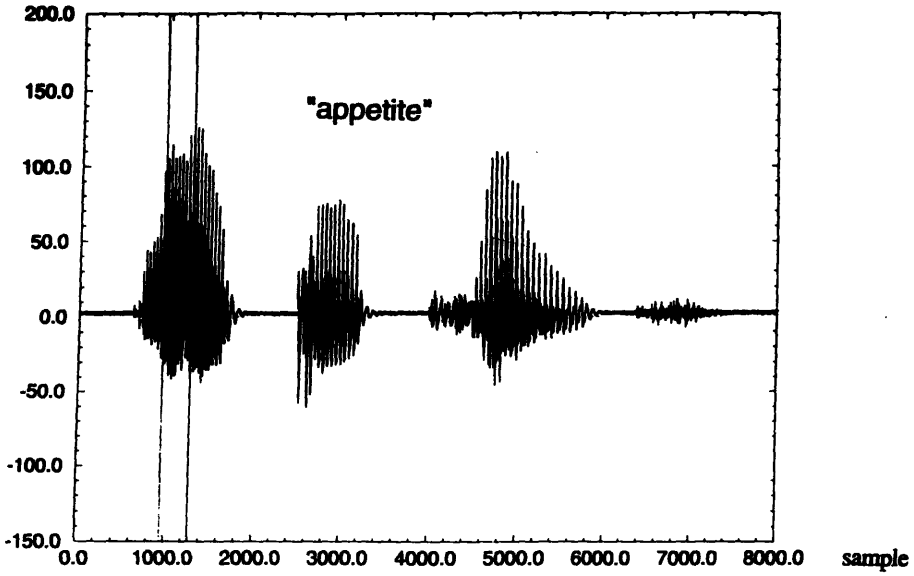
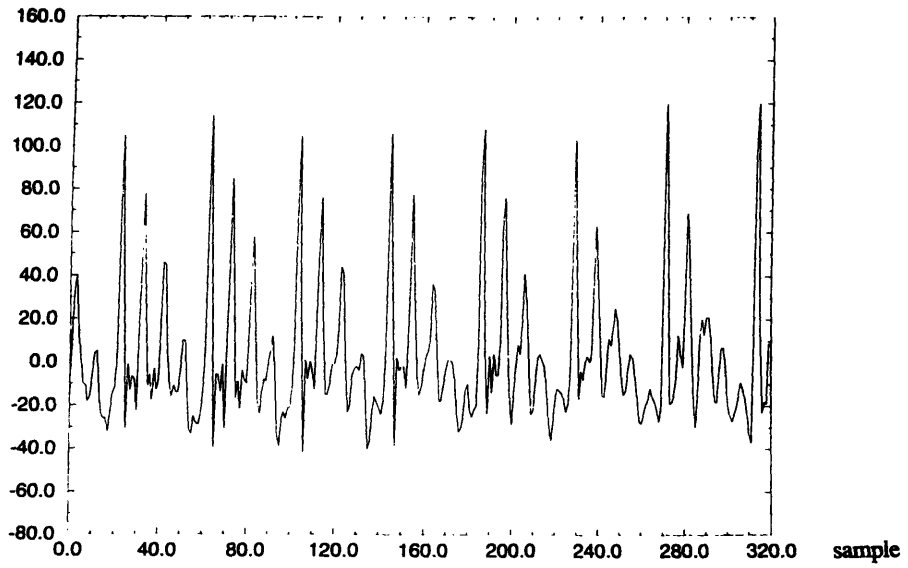


Figure 5.2-1: Human speech production system. (After Flanagan [88])

Figure 5.2-2: Opening and closing of the human vocal cords. (After Fletcher [86])



(a)



(b)

Figure 5.2-3: Word and voiced speech frame.
(a) Time plot of the word "appetite" sampled at 8 kHz.
(b) 320 samples (40 ms) of the voiced sound \a\ at the front of "appetite."Also represented as the samples between the two vertical lines in Fig. 5.2-3a.

Over the years many strategies have been proposed to tackle this problem because accurate pitch detection poses many challenges:

- Vocal cord vibration is not exactly periodic, especially at the beginning and end of voiced sounds [94].
- Pitch can be corrupted by the background resonances of the vocal tract, i.e., formants.
- Variations in pitch can come from many factors, e.g., a speaker's speaking style, his language, or his vocal tract characteristics.
- Pitch period has a very large dynamic range and varies from 1.25 ms to 40 ms ([85], [94]).
- Pitch can be corrupted by background noise which develops pitch structure when the noise level is high enough.
- Pitch detection is a non-stationary signal analysis problem, and a trade-off exists between the performance and complexity of any algorithm.

Hermes [89] has classified all pitch-determination algorithms (PDA's) into three generations:

- First generation: Estimate the first harmonic or the period in voiced speech.
- Second generation: Based on models and theories of pitch perception.
- Third generation: Based on pitch perception and the human auditory system.

The type under consideration in this work is a mixture of the first and third generations.

Two first generation methods based on the cepstrum [79] and autocorrelation [68], respectively, have remained popular over the years. The cepstrum method entails applying the inverse Fourier transform to the log-power spectrum of a signal to separate the spectral envelope due the vocal tract from the source pitch pulses. The autocorrelation method, on the other hand, entails evaluating the autocorrelation function of the input waveform, and may include some preprocessing such as center or peak clipping for spectrum flattening [94]. The latter scheme and its performance are considered here for the purpose of a better evaluation of and comparison with the methods proposed in the next section. The testing speech data consist of:

- Clean speech: 320 samples plotted in Fig. 5.2-3b.
- Noisy speech: Obtained by adding to the clean speech appropriately scaled white Gaussian noise to achieve the following seven different signal-to-ratios (SNR's): 30 dB, 25 dB, 20 dB, 15 dB, 10 dB, 5 dB, and 4 dB.

Figs. 5.2-5a through 5.2-5h show the normalized autocorrelation coefficients for the clean and noisy testing data (Figs. 5.2-4a through 5.2-4h). Performance is judged by how obvious and detectable the pitch peaks are. From the figures, we see that the performance is poor because the peaks are hardly distinguishable from the vocal tract effects (formants). The situation worsens in the presence of noise.

To alleviate the problematic role played by the formants, we use center-clipping, in which a certain percentage of the maximum amplitude is subtracted from all the samples before autocorrelation is carried out. If $x(n)$ denotes the sampled input signal, then the center-clipped output, $y(n)$, is defined as [15]

$$y(n) = C[x(n)],$$

where the $C[\bullet]$ operation is shown in Fig. 5.2-6, and C_L is set equal to 30% of the maximum amplitude.

Autocorrelation coefficients for the center-clipped versions of Fig. 5.2-4a through Fig. 5.2-4h are plotted from Fig. 5.2-7a to Fig. 5.2-7h. Although better than their unclipped counterparts, the results still reveal a disruptive role played by the peaks due to the formants. To compensate for this, the peak-picking algorithm will have to be very complex. A scheme to tip the trade-off balance in favor of a simpler peak-picking algorithm is proposed and tested in the next two sections.

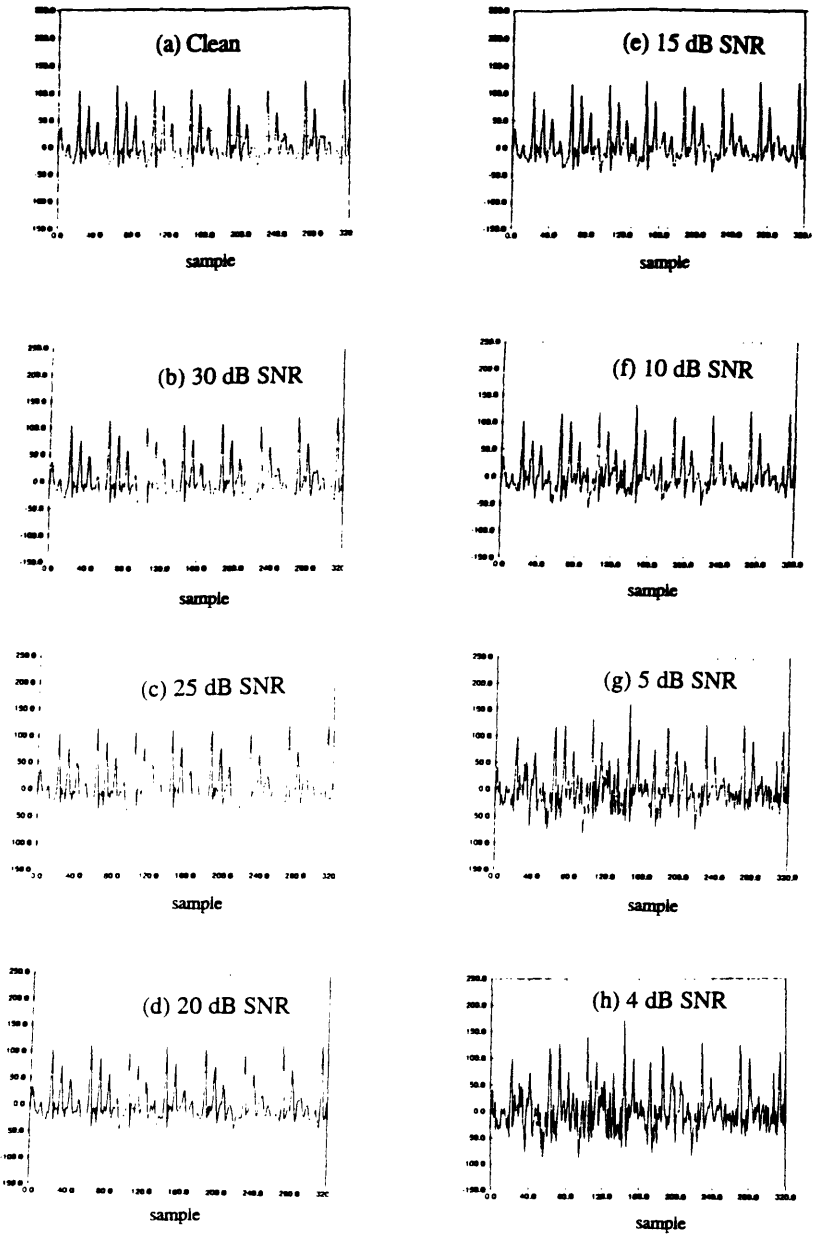
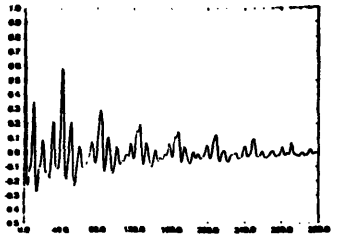
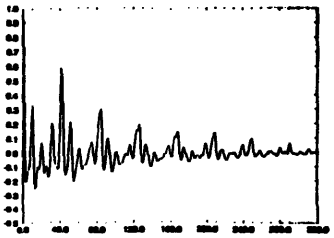


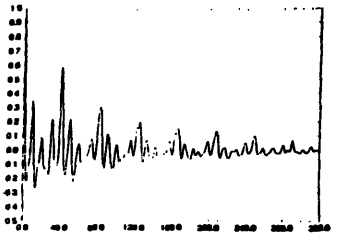
Figure 5.2-4: Clean and noisy test speech data.
(a) Clean, (b) 30 dB SNR, (c) 25 dB SNR, (d) 20 dB SNR
(e) 15 dB SNR, (f) 10 dB SNR, (g) 5 dB SNR, (h) 4 dB SNR



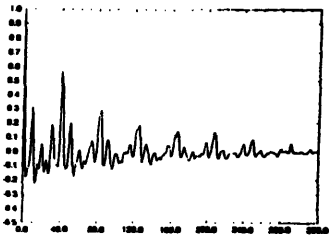
(a) Coefficients for Clean.



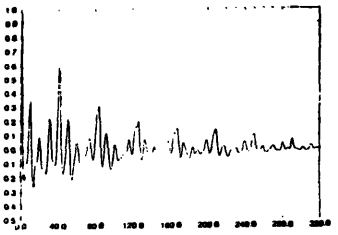
(e) Coefficients for 15 dB SNR.



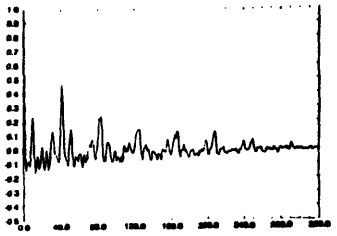
(b) Coefficients for 30 dB SNR.



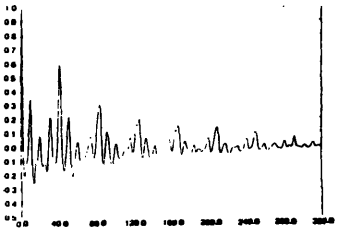
(f) Coefficients for 10 dB SNR.



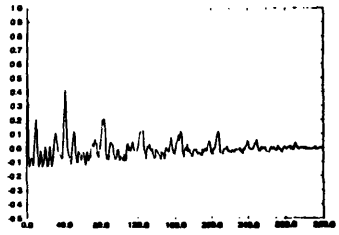
(c) Coefficients for 25 dB SNR.



(g) Coefficients for 5 dB SNR.



(d) Coefficients for 20 dB SNR.



(h) Coefficients for 4 dB SNR.

Figure 5.2-5: Normalized autocorrelation coefficients for clean and noisy test speech samples. For
(a) Clean, (b) 30 dB SNR, (c) 25 dB SNR, (d) 20 dB SNR
(e) 15 dB SNR, (f) 10 dB SNR, (g) 5 dB SNR, (h) 4 dB SNR

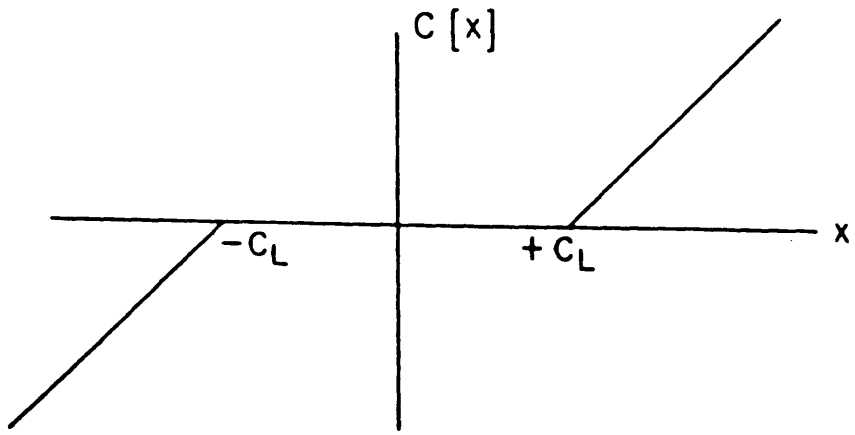
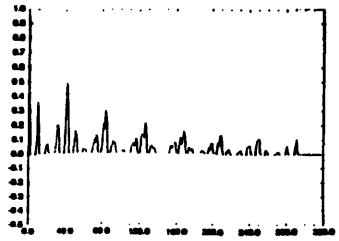
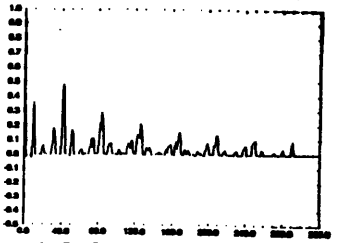


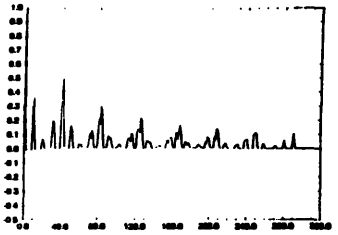
Figure 5.2-6: Center clipping operation, C[]. (After Rabiner and Schafer [15])



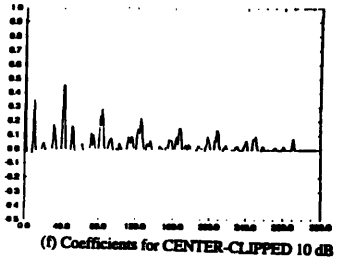
(a) Coefficients for CENTER-CLIPPED Clean.



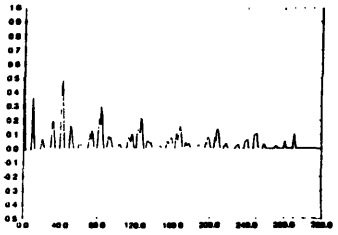
(e) Coefficients for CENTER-CLIPPED 15 dB SNR.



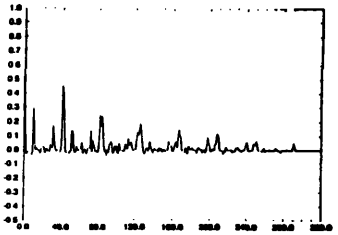
(b) Coefficients for CENTER-CLIPPED 30 dB SNR.



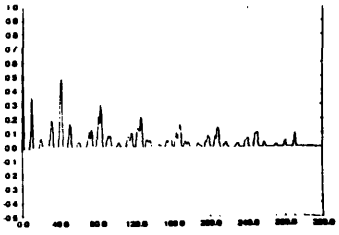
(f) Coefficients for CENTER-CLIPPED 10 dB SNR.



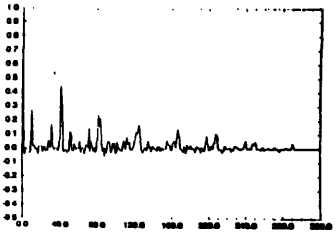
(c) Coefficients for CENTER-CLIPPED 25 dB SNR.



(g) Coefficients for CENTER-CLIPPED 5 dB SNR.



(d) Coefficients for CENTER-CLIPPED 20 dB SNR.



(h) Coefficients for CENTER-CLIPPED 4 dB SNR.

Figure 5.2-7: Normalized autocorrelation coefficients for CENTER-CLIPPED clean and noisy test samples. For
(a) Clean, (b) 30 dB SNR, (c) 25 dB SNR, (d) 20 dB SNR,
(e) 15 dB SNR, (f) 10 dB SNR, (g) 5 dB SNR, (h) 4 dB SNR

5.3 Pitch Detection with Basic Spline Wavelet Transform

One impetus to use the wavelet transform to detect pitch is based on a finding by Mallat and Zhong [35] that non-stationarities in a signal show up as wavelet transform maxima. In our application, the non-stationary instant of glottal closure can be detected, and the time between two such consecutive events, or two consecutive maxima in the wavelet transform coefficients, is the pitch period.

Another impetus to use the transform is based on its capability to model the human visual and auditory systems. The human eye is still one of the most reliable estimators of the pitch period in a time-plotted speech signal. The human ear is also an excellent pitch detector, especially in the presence of noise, for even in a very noisy setting, our ears still enable us to engage in an intelligible conversation. Since the transform is a good model of main characteristics of these two systems, it has great potential in this application.

Using Mallat's fast wavelet algorithm, Kadambe and Boudreaux-Bartels [85] have studied the performance of the wavelet transform in pitch detection and found that it performed much better than the conventional cepstrum and autocorrelation methods. The wavelet they used is the same one used by Mallat and Zhong for edge type discrimination [35], namely the first derivative of a smoothing function. Fig. 5.3-1 shows the smoothing function, a cubic B-spline, and the quadratic spline wavelet. (An important observation to be made here is that the wavelet can be approximated by two shifted and scaled cubic B-splines. The performance of such a B-spline approximated wavelet is evaluated in the next section.)

The use of the first derivative of a smoothing function as a wavelet has a very interesting interpretation. What we are doing here is looking at the first derivative of a

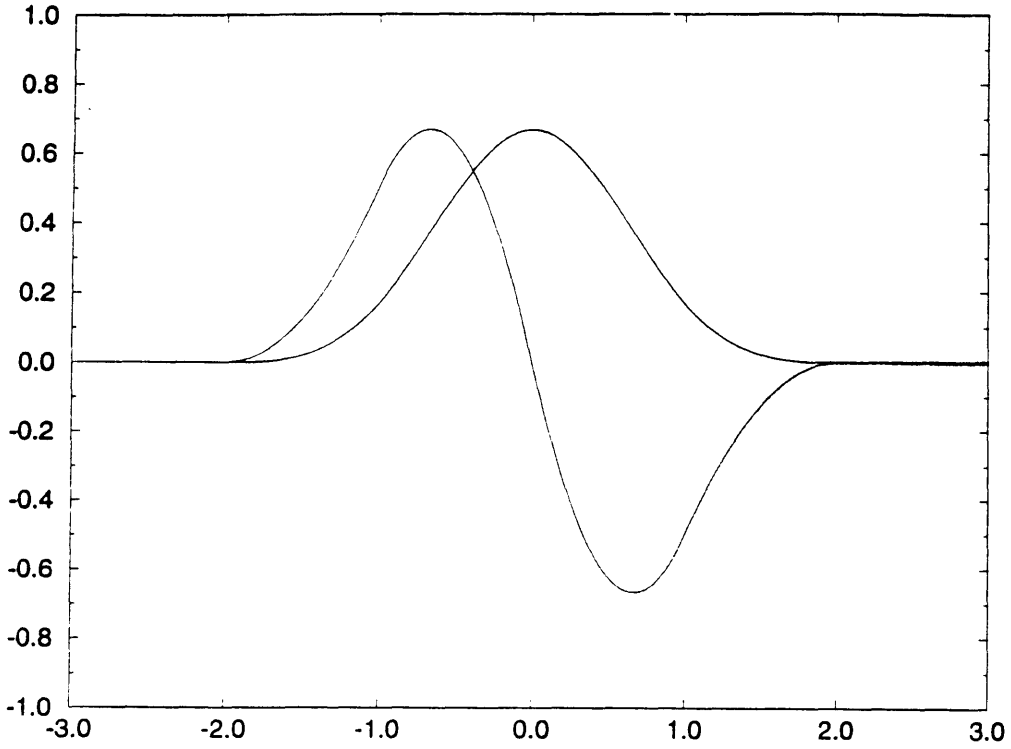


Figure 5.3-1: The smoothing cubic B-spline function and its first derivative, the quadratic spline.

smoothed version of the input signal so that any sudden perturbations will give rise to transform maxima. This approach can be very useful in signal analysis, especially in pitch detection. However, the type of wavelet one should choose for a certain task is still very nebulous. If the rule is that the wavelet must be the first derivative of a smoothing function, then any arbitrarily shaped degree n polynomial spline can qualify as a wavelet because it is always a derivative of a smoothing spline of degree $n+1$. In addition, we need to address how high n should be, for we see in Fig. 4.3-1 that functional support increases along with n. In an attempt to clarify these and other ambiguities, we use the basic spline wavelet transform of equation (4.4-15) and experiment with various wavelet designs. The major issues under consideration are:

- Wavelet shape, parameter $q[k]$ in equation (4.4-15).
- Order of approximating B-splines, parameter n_2.
- Best scale to pick out the pitch peaks, parameter m.

We will also evaluate the transform's performance in noisy and multi-speaker test data, and see what improvements can be made.

5.4 Experimental Results and Discussion

5.4.1 Wavelet Shape

In this set of experiments we seek to find the general shape of a wavelet most suitable for pitch detection. The testing data is the frame of 320 clean speech samples shown in Fig. 5.2-3b. Transform coefficients are obtained by implementing the BSWT of equation (4.4-15) on a digital computer. Wavelets are designed with cubic B-splines (i.e., $n_2 = 3$). After experimenting with the four dyadic viewing scales (parameter m) of 2, 4, 8, and 16, we find that $m = 8$ furnishes the best results.

Eight wavelets are designed and their performances are evaluated. Table 5.4.1-1 (at the end of this section) shows the *q[k]* coefficients for each wavelet, and Figs. 5.4.1-1 to 5.4.1-8 (also at the end of this section) show both the wavelet shapes (part (a)) and the BSWT coefficients (part (b)). Of the eight wavelets, only W1, W2, and W4 are admissible. The rest can be viewed as matched filters [90] and are used to determine how well the input signal can be decomposed in terms of these "small waves." (Hereafter, we use the term *wavelet* to denote these "small waves" whether they are admissible or not.)

To check the performance of this approach, we obtain the actual samples of pitch onset (i.e., glottal closure) by hand, and plot them as vertical lines in part (b) of each figure. Pitch detection suitability is evaluated according to two criteria:

(1) How obvious are the pitch peaks (i.e., the local maxima in the BSWT coefficients).

(2) How well do these peaks coincide with the vertical lines, or actual samples of pitch onset.

An examination of part (b) of each of the figures shows that wavelet W1 is the most suitable. The performance of the other wavelets furnishes us with some insight into factors that warrant careful consideration:

- Symmetric vs. Antisymmetric: The simplest antisymmetric wavelet, W1, performs much better than any of the symmetric or antisymmetric wavelets with larger support. The shape of the symmetric wavelets are too restrictive: that is, there are too many lobes and/or the support is too big. Consequently, false pitch peaks contributed by background formants are picked up. The restrictive shape has caused compensative irregularities (false local maxima) in the transform coefficients.
- Support: A comparison of the results of W1 with W2 and of W4 with W5 tells us that as the wavelet support increases, the shape becomes more restrictive. In cases in which the shape is too restrictive, not only are there false pitch peaks, but they are larger than the actual ones. The important role of support is further studied in Section 5.4.3.

- Matched filter approached: The shapes of W6, W7, and W8 are designed so as to model the onset of a pitch pulse. But they performed very poorly. In particular, the lack of a negative lobe in W8 causes false peaks to be more dominant than the actual ones.

In the experiments of the following sections, we focus our attention on wavelet W1 and others similar to it.

Table 5.4.1-1: The *q[k]* coefficients of wavelets W1-W8.

W1*: *q[-1] = 1.0*
q[0] = -1.0

W2*: *q[-2] = 1.0*
q[-1] = -1.0
q[0] = 1.0
q[1] = -1.0

W3: *q[-1] = -1.0*
q[0] = 1.0
q[1] = -1.0

W4*: *q[-1] = -1.0*
q[0] = 2.0
q[1] = 1.0

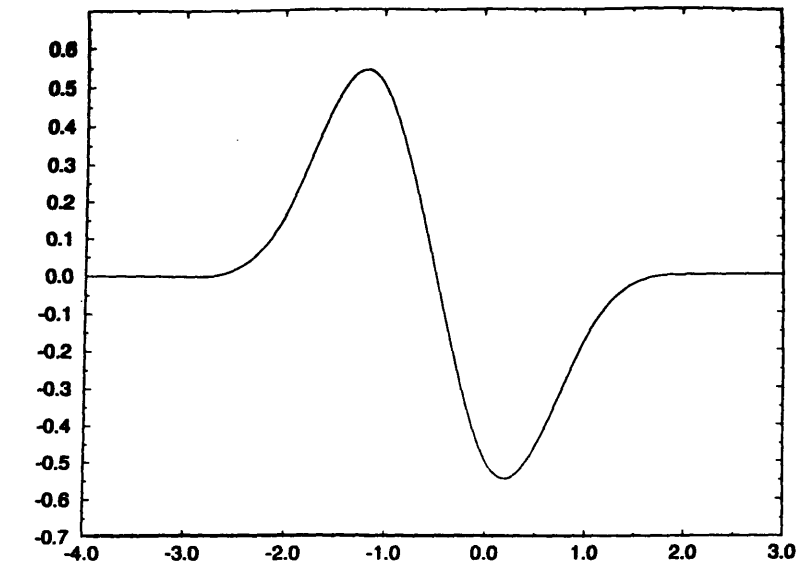
W5: *q[-2] = 1.0*
q[-1] = -2.0
q[0] = 3.0
q[1] = -2.0
q[2] = 1.0

W6: *q[-2] = 1.0*
q[-1] = -1.0
q[0] = 2.0
q[1] = -2.0
q[2] = 3.0

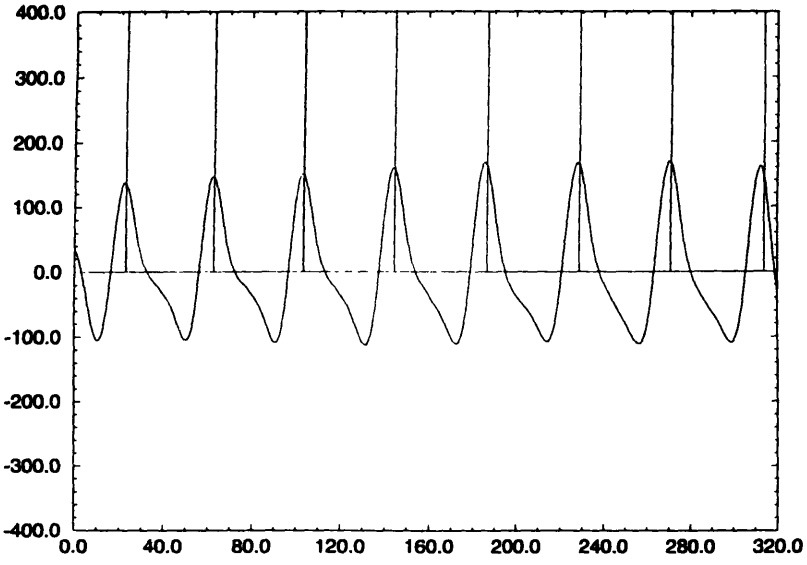
W7: *q[-2] = 3.0*
q[-1] = -2.0
q[0] = 2.0
q[1] = -1.0
q[2] = 1.0

W8: *q[-2] = 3.0*
q[-1] = -0.25
q[0] = 2.0
q[1] = -0.125
q[2] = 1.0

*- admissible

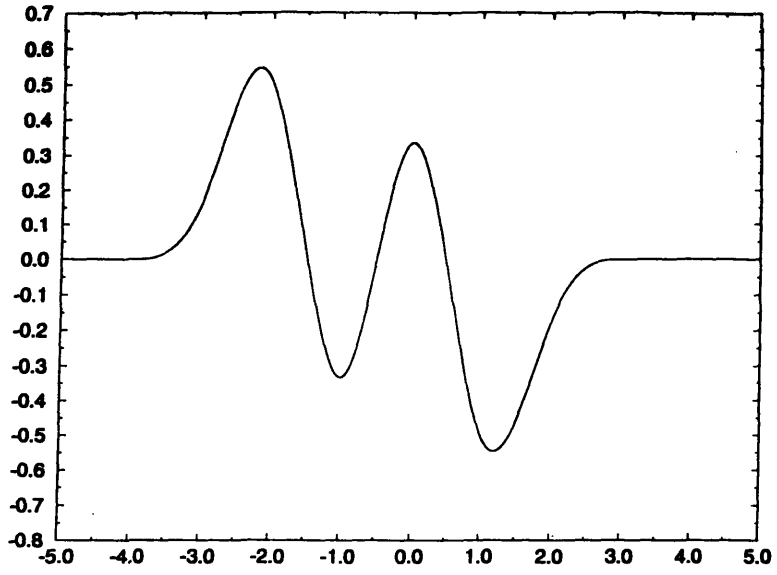


(a) W1 shape

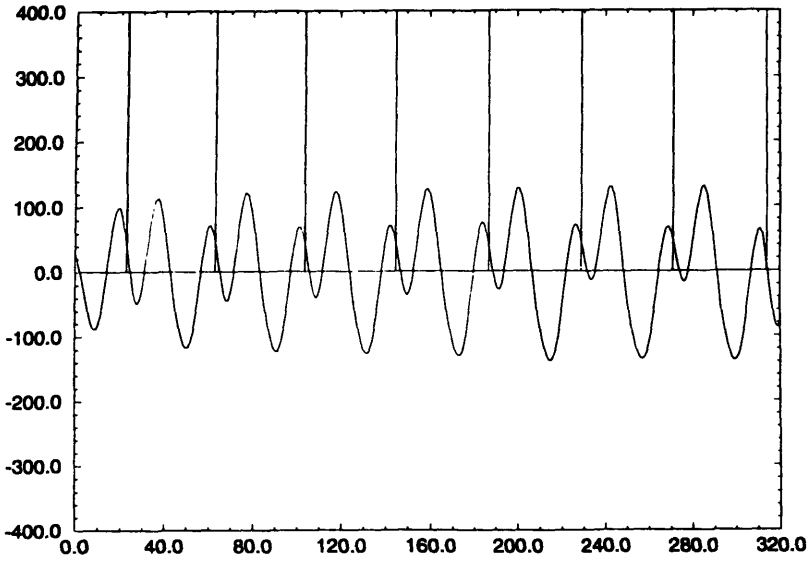


(b) W1 BSWT coefficients

Figure 5.4.1-1: Wavelet W1: shape and BSWT coefficients.
(a) Shape
(b) BSWT coefficients.

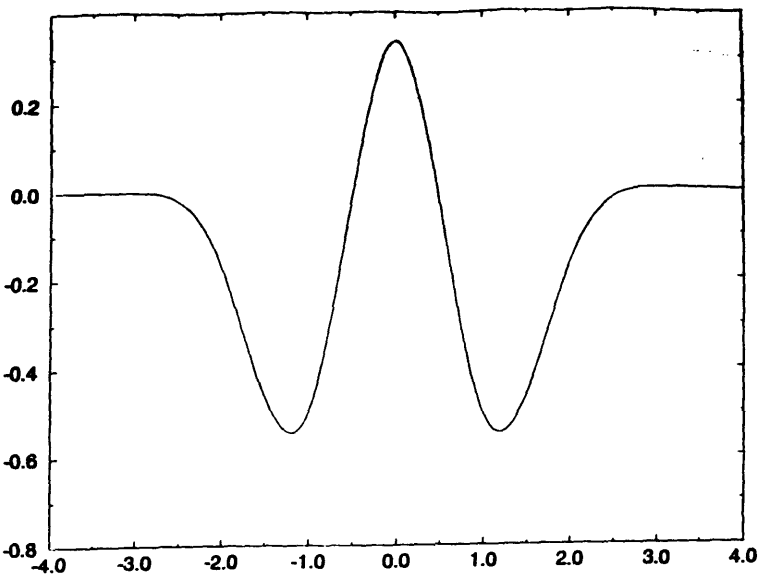


(a) W2 shape

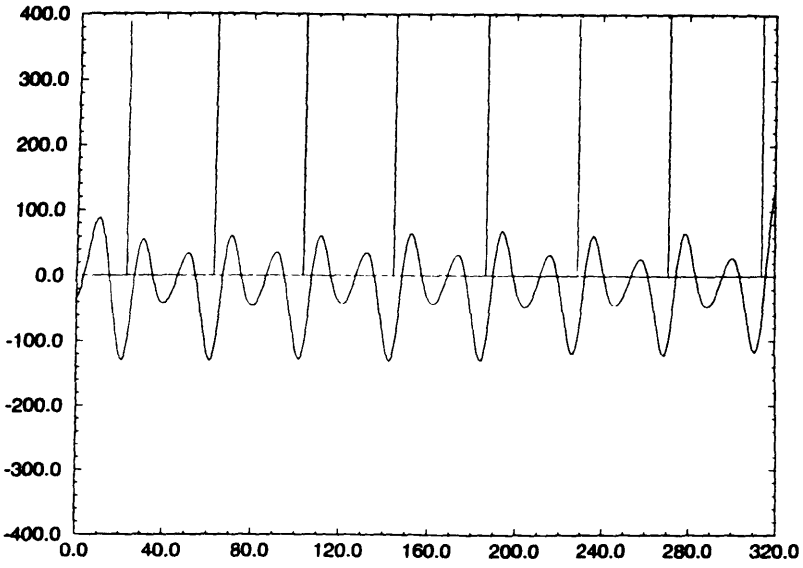


(b) W2 BSWT coefficients

Figure 5.4.1-2: Wavelet W2: shape and BSWT coefficients.
(a) Shape
(b) BSWT coefficients

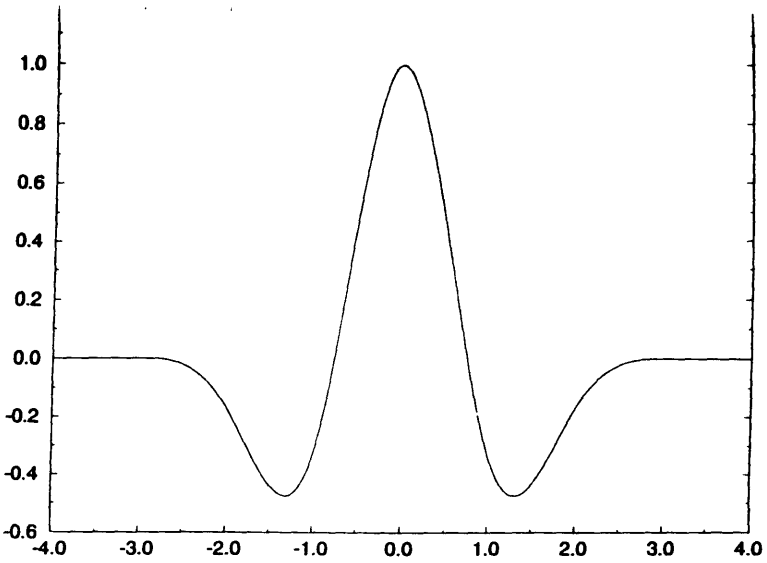


(a) W3 shape

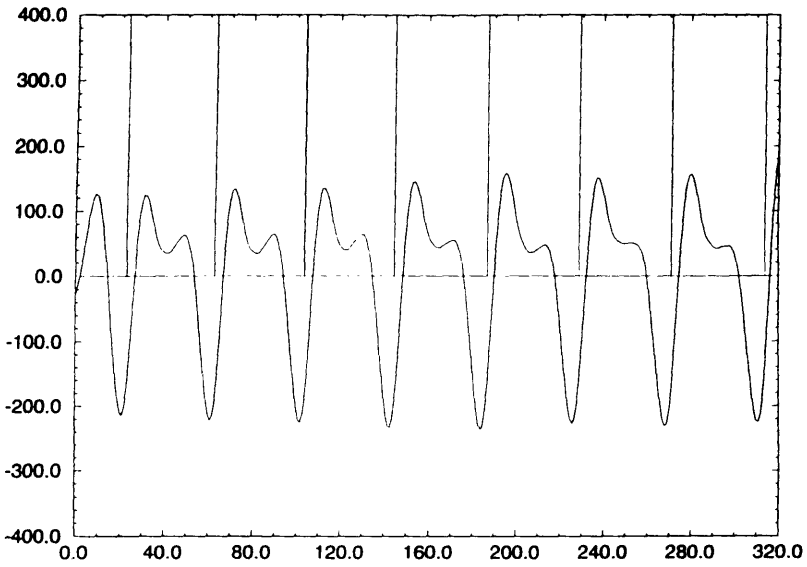


(b) W3 BSWT coefficients

Figure 5.4.1-3: Wavelet W3: shape and BSWT coefficients.
(a) Shape
(b) BSWT coefficients

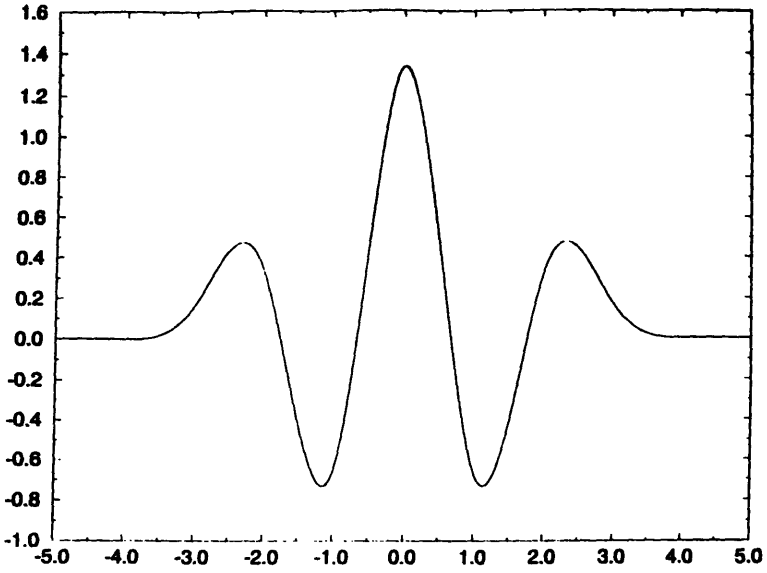


(a) W4 shape

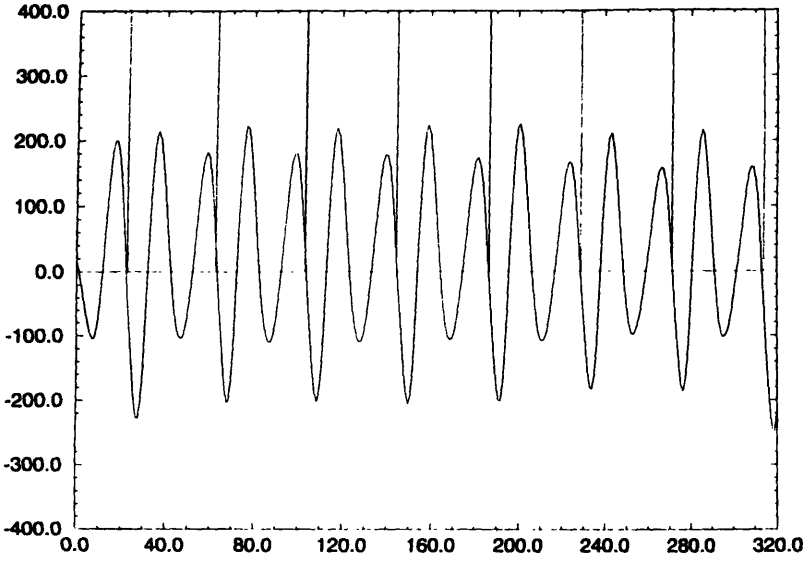


(b) W4 BSWT coefficients

Figure 5.4.1-4: Wavelet W4: shape and BSWT coefficients.
(a) Shape
(b) BSWT coefficients

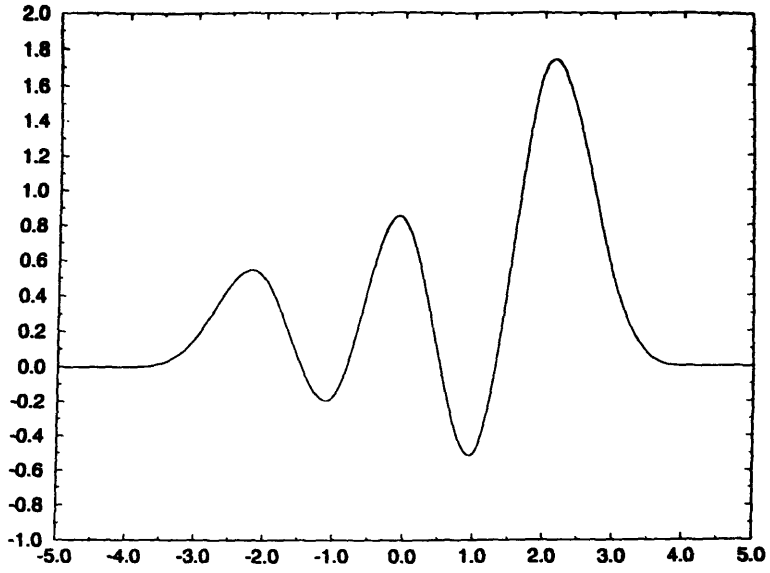


(a) W5 shape

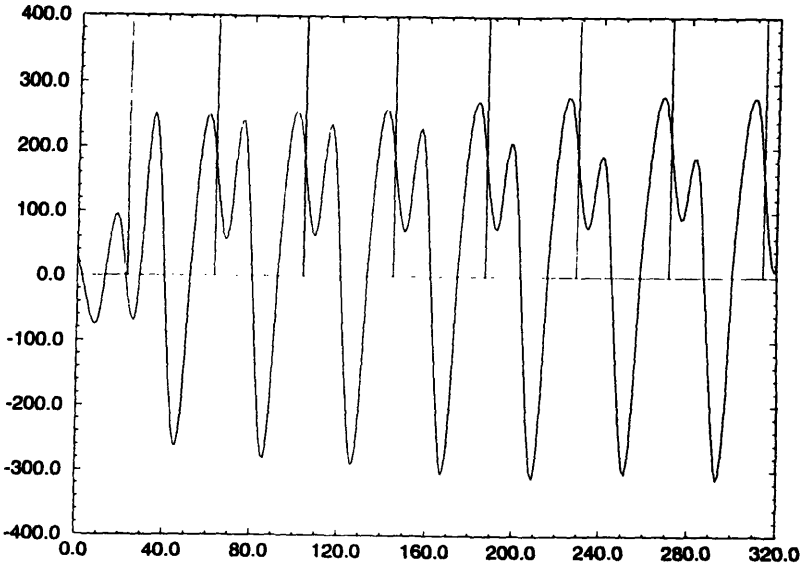


(b) W5 BSWT coefficients

Figure 5.4.1-5: Wavelet W5: shape and BSWT coefficients.
(a) Shape
(b) BSWT coefficients

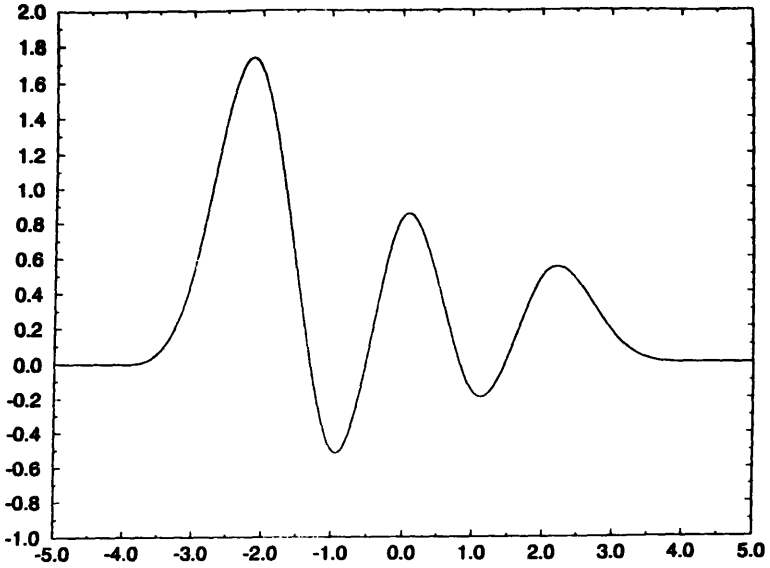


(a) W6 shape

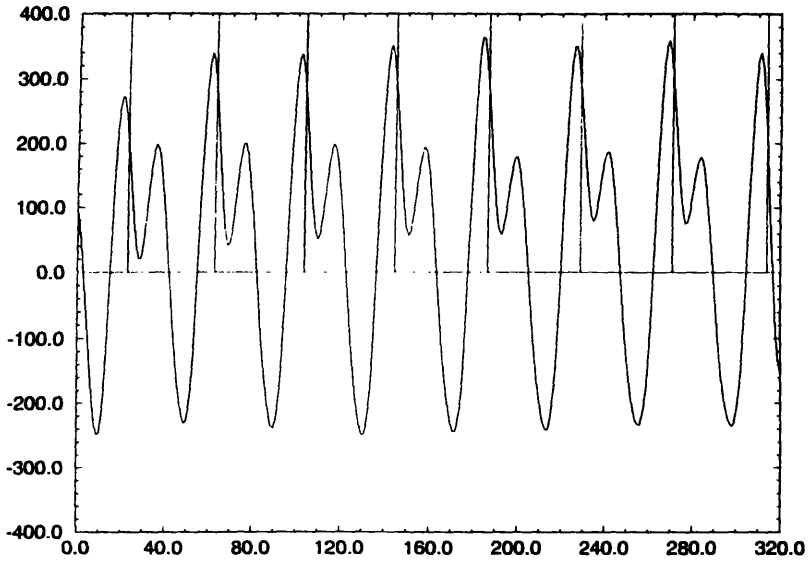


(b) W6 BSWT coefficients

Figure 5.4.1-6: Wavelet W6: shape and BSWT coefficients.
(a) Shape
(b) BSWT coefficients

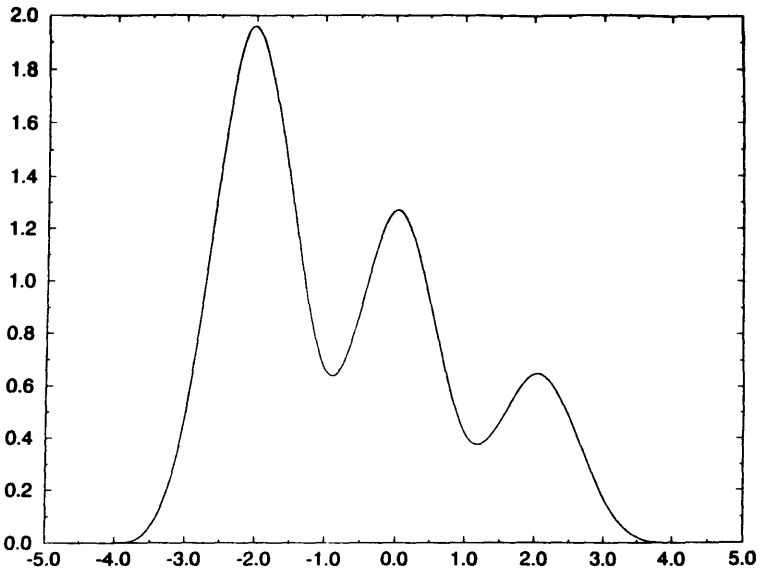


(a) W7 shape

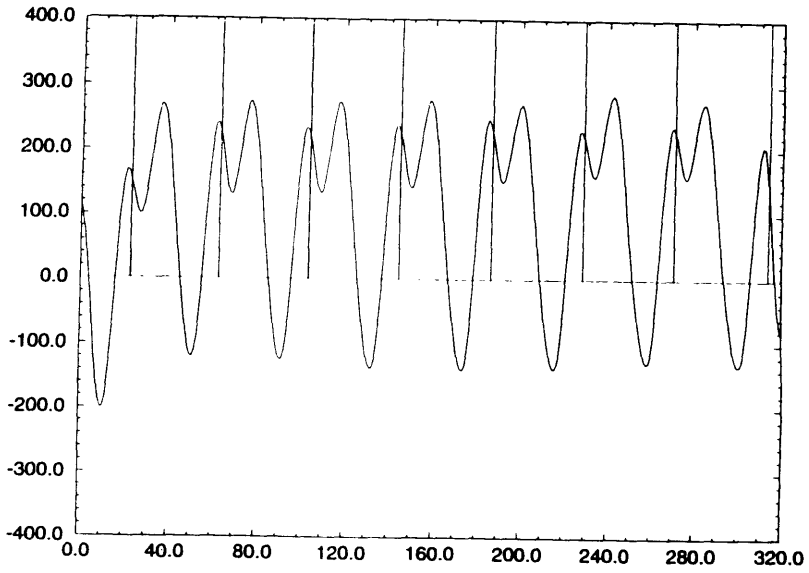


(b) W7 BSWT coefficients

Figure 5.4.1-7: Wavelet W7: shape and BSWT coefficients.
(a) Shape
(b) BSWT coefficients



(a) W8 shape



(b) W8 BSWT coefficients

Figure 5.4.1-8: Wavelet W8: shape and BSWT coefficients.
(a) Shape
(b) BSWT coefficients

5.4.2 Approximating Spline Order and Viewing Scale

To find the best spline order to approximate a wavelet and the best viewing scale, we evaluate the transform at the orders (parameter n_2) of 1st, 3rd, 5th, and 7th, and at the dyadic viewing scales of (parameter m) 2, 4, 8, and 16. Fig. 5.4.2-1 shows the W1 wavelet approximated with four different orders of B-splines. Different viewing scales for each of the spline orders are shown from Fig. 5.4.2-2 to Fig. 5.4.2-5.

From the experimental results, the best scale is $m = 8$. Viewed at the lowest scale ($m = 2$), much of the signal details is preserved: the BSWT coefficients look very much like the original speech waveform. At the highest scale ($m = 16$), the details are smoothed out. At $m = 8$, the dominant feature (i.e., pitch onset) is still retained, while the unimportant, and disruptive, details are smoothed out significantly.

As for the best approximating order at $m = 8$, the 3rd and 7th orders have very similar performances. However, taking into account that the 7th-order splines are much more computationally demanding, we conclude that the 3rd-order splines are the best. Hereafter, all experimental wavelets will be approximated with cubic B-splines and the BSWT viewed at a scale of $m = 8$.

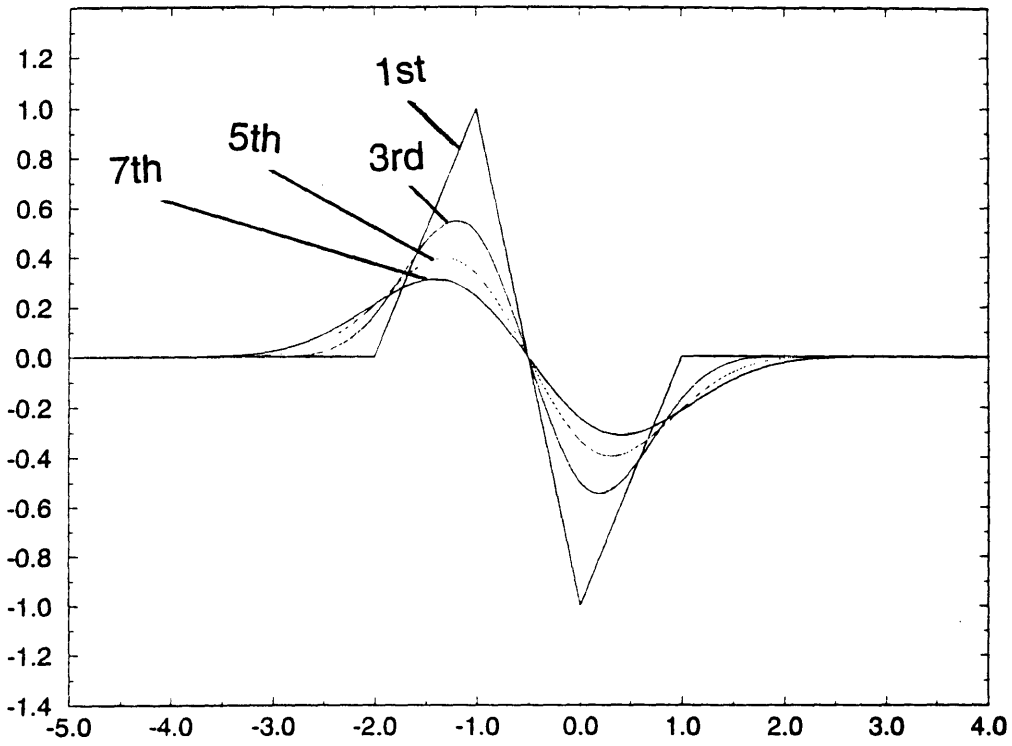


Figure 5.4.2-1: Wavelet W1 approximated with four different orders of B-splines.

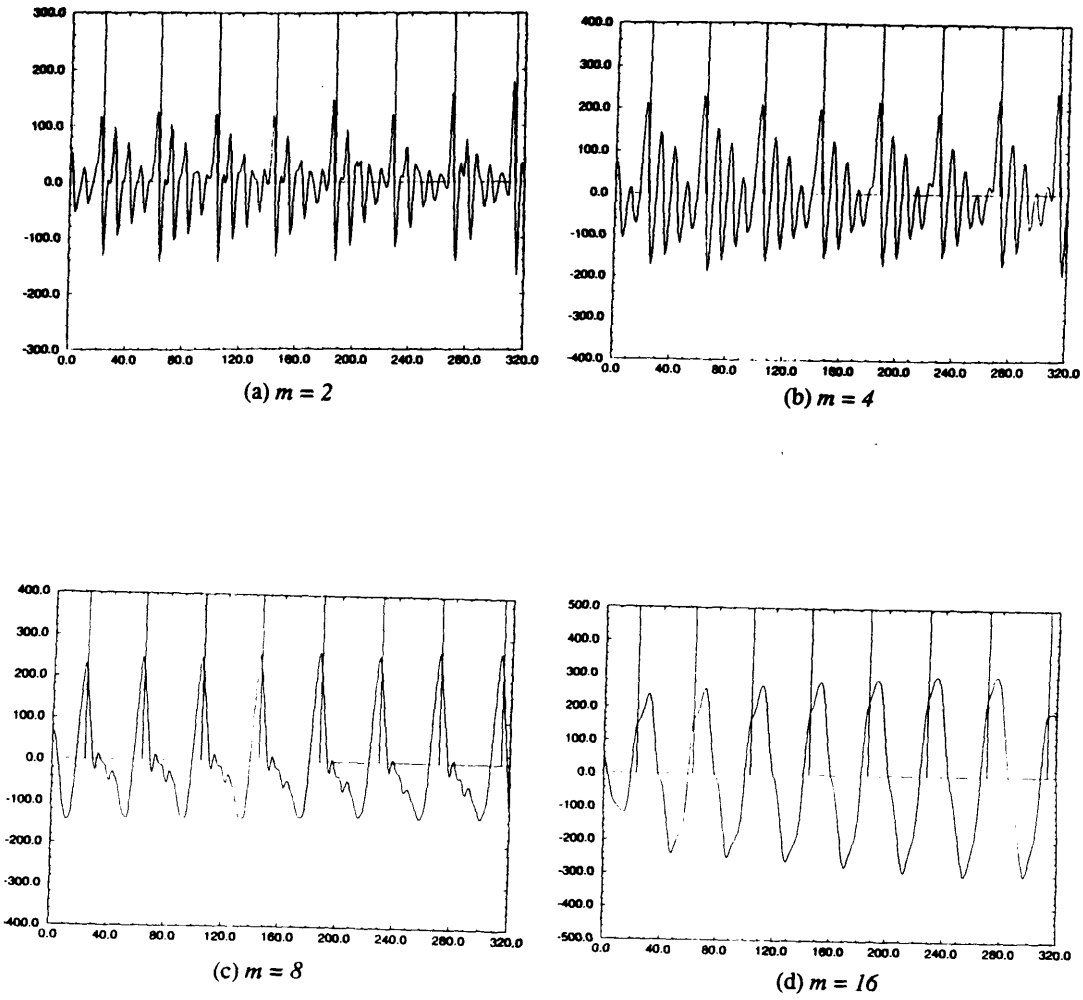
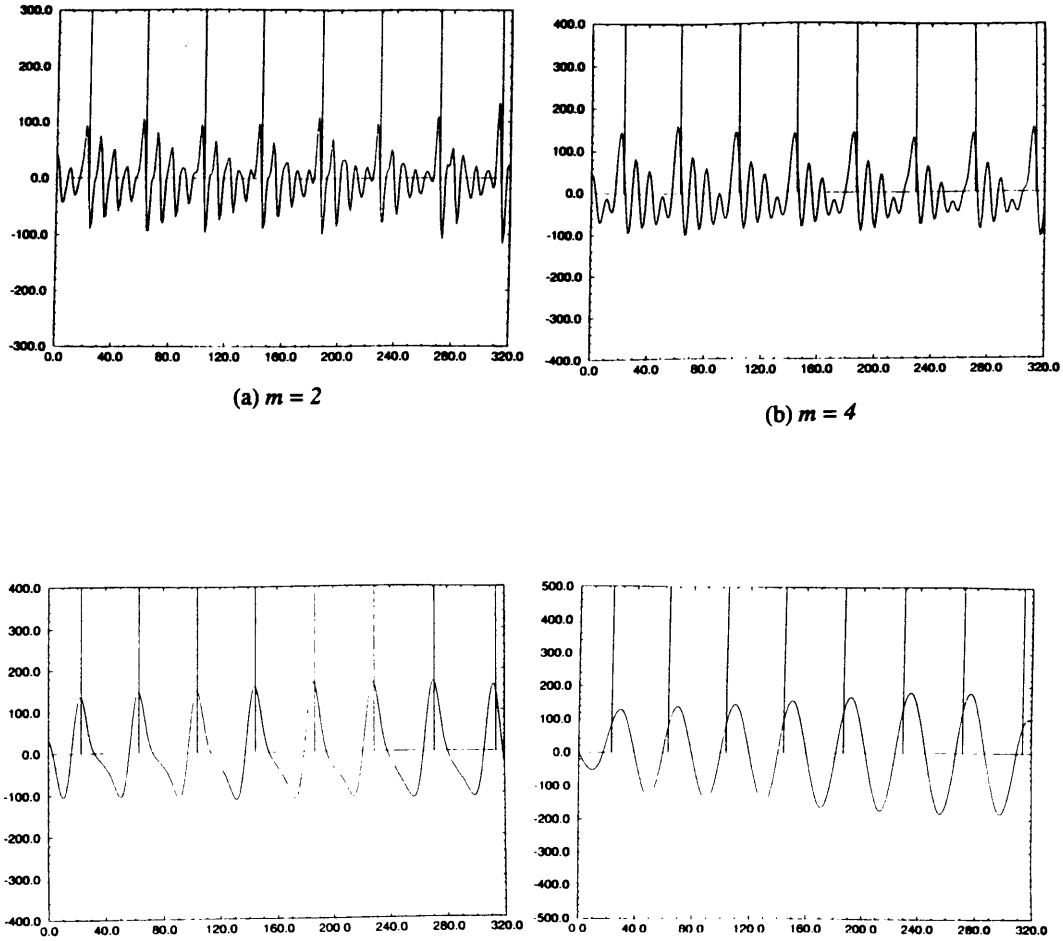


Figure 5.4.2-2: BSWT coefficients based on W1 approximated with 1st-order B-splines and viewed at four different scales. At scale
(a) $m = 2$, (b) $m = 4$, (c) $m = 8$, (d) $m = 16$.



(a) $m = 2$ (b) $m = 4$

(c) $m = 8$ (d) $m = 16$

Figure 5.4.2-3: BSWT coefficients based on W1 approximated with 3rd-order B-splines and viewed at four different scales. At scale
(a) $m = 2$, (b) $m = 4$, (c) $m = 8$, (d) $m = 16$.

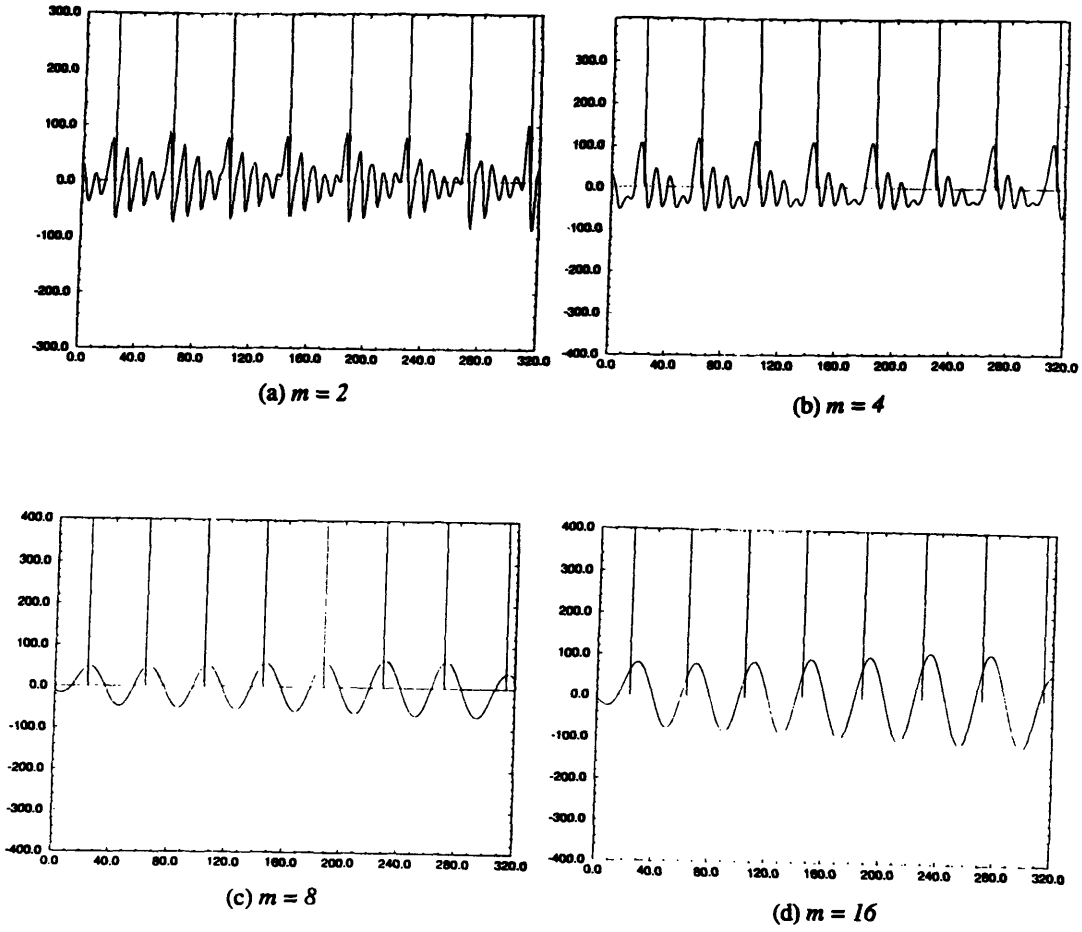


Figure 5.4.2-4: BSWT coefficients based on W1 approximated with 5th-order B-splines and viewed at four different scales. At scale
(a) $m = 2$, (b) $m = 4$, (c) $m = 8$, (d) $m = 16$.

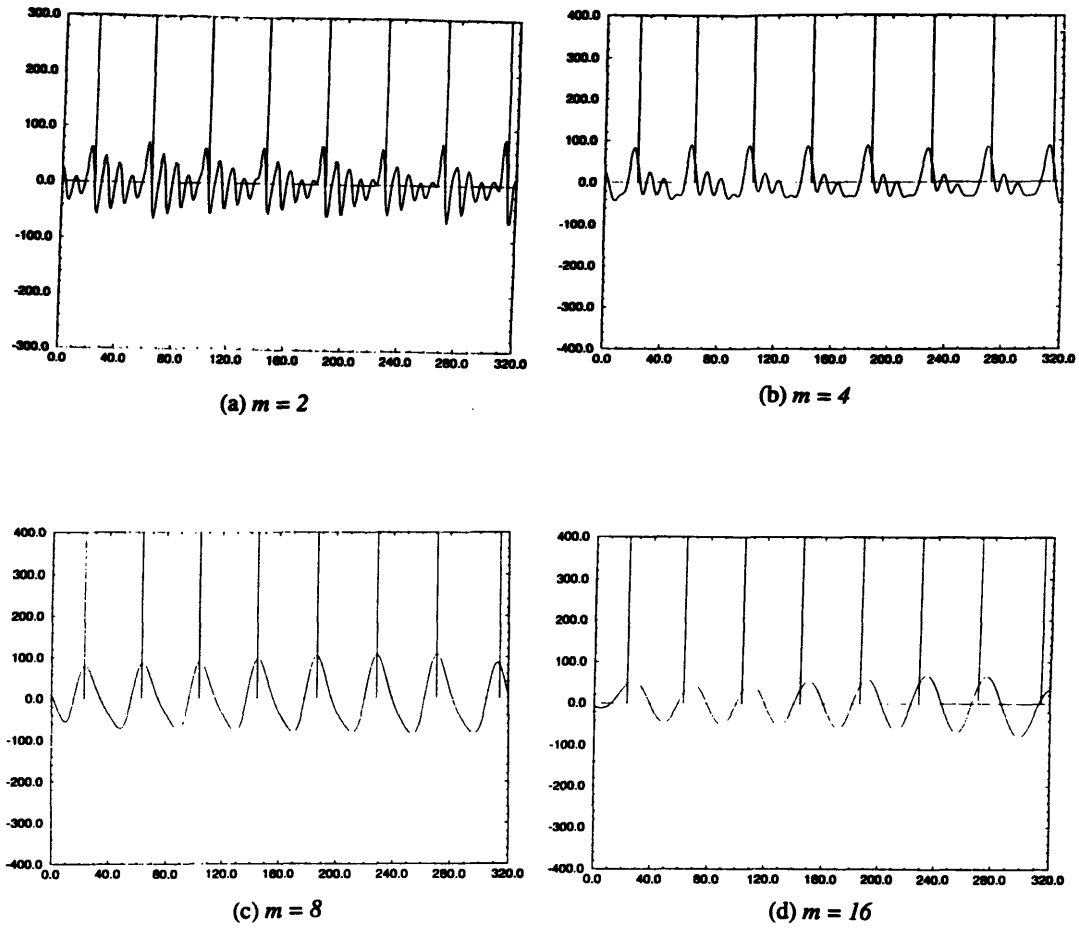


Figure 5.4.2-5: BSWT coefficients based on W1 approximated with 7th-order B-splines and viewed at four different scales. At scale
(a) $m = 2$, (b) $m = 4$, (c) $m = 8$, (d) $m = 16$.

5.4.3 Wavelet Support

The importance of wavelet support, alluded to in Section 5.4.1, is further examined in the following experiments.

Five wavelets, WA-WE, with varying supports are designed. Their *q[k]* parameters can be found in Table 5.4.3-1; their shapes approximated with cubic B-splines and their BSWT coefficients viewed at a scale of $m = 8$ for the clean speech test samples (Fig. 5.2-3b) are shown in Figs. 5.4.3-2 to 5.4.3-6. Fig. 5.4.3-1 shows a rescaled version of the BSWT coefficients of W1 for the purpose of a better comparison with the newly designed wavelets.

Comparing the results of W1 (Fig. 5.4.3-1) with those of WA (Fig. 5.4.3-2) and WB (Fig. 5.4.3-3b), we see that as the support increases, so do the transform amplitudes, and the peaks are shifted away from the actual pitch pulses. Because they are also sources of non-stationarities, the background formants are inadvertently picked as pitch onsets. A compromise occurs in which the peaks are shifted toward these formants. To see what this means graphically, let us compare the results of W1 (Fig. 5.4.3-1) with those of WC (Fig. 5.4.3-4). Though they are very similar, the pitch peaks of WC do have a slight shift.

This shifting becomes very evident in the results of wavelets whose support is too big and shape too inflexible (e.g., Fig. 5.4.3-5 and Fig. 5.4.3-6).

In summary, this set of experimental data shows that support plays a very important role in the accurate detection of the instant of pitch onset. Minimal support is essential.

Table 5.4.3-1: The *q[k]* coefficients of wavelets WA-WE with varying supports.

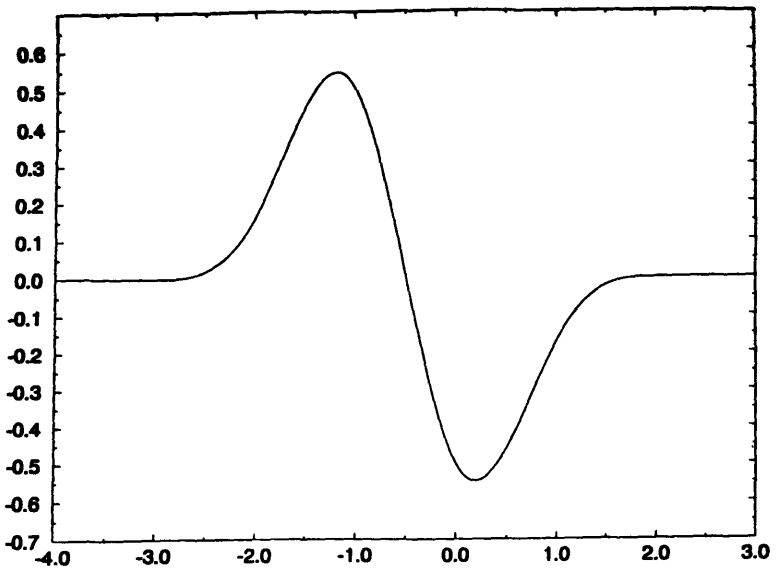
WA: *q[-1] = 1.0*
q[0] = 0.0
q[-1] = -1.0

WB: *q[-3] = 1.0*
q[-2] = 2.0
q[-1] = 1.0
q[0] = 0.0
q[1] = -1.0
q[2] = -2.0
q[3] = -1.0

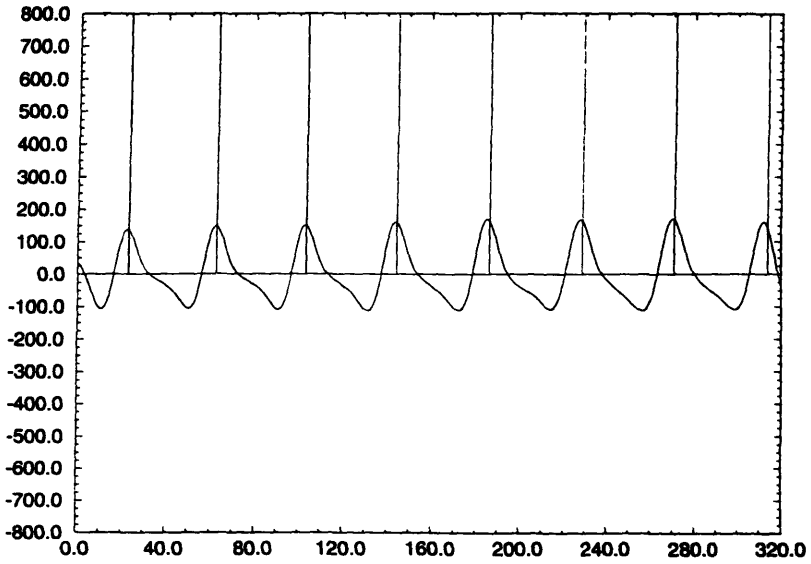
WC: *q[-3] = 1.0*
q[-2] = 0.0
q[-1] = -1.0
q[0] = 0.0
q[1] = 1.0
q[2] = 0.0
q[3] = -1.0

WD: *q[-7] = 1.0*
q[-6] = 0.0
q[-5] = -1.0
q[-4] = 0.0
q[-3] = 1.0
q[-2] = 0.0
q[-1] = -1.0
q[0] = 0.0
q[1] = 1.0
q[2] = 0.0
q[3] = -1.0
q[4] = 0.0
q[5] = 1.0
q[6] = 0.0
q[7] = -1.0

WE: *q[-15] = 1.0*
q[-14] = 0.0
q[-13] = -1.0
q[-12] = 0.0
q[-11] = 1.0
q[-10] = 0.0
q[-9] = -1.0
q[-8] = 0.0
q[-7] = 1.0
q[-6] = 0.0
q[-5] = -1.0
q[-4] = 0.0
q[-3] = 1.0
q[-2] = 0.0
q[-1] = -1.0
q[0] = 0.0
q[1] = 1.0
q[2] = 0.0
q[3] = -1.0
q[4] = 0.0
q[5] = 1.0
q[6] = 0.0
q[7] = -1.0
q[8] = 0.0
q[9] = 1.0
q[10] = 0.0
q[11] = -1.0
q[12] = 0.0
q[13] = 1.0
q[14] = 0.0
q[15] = -1.0

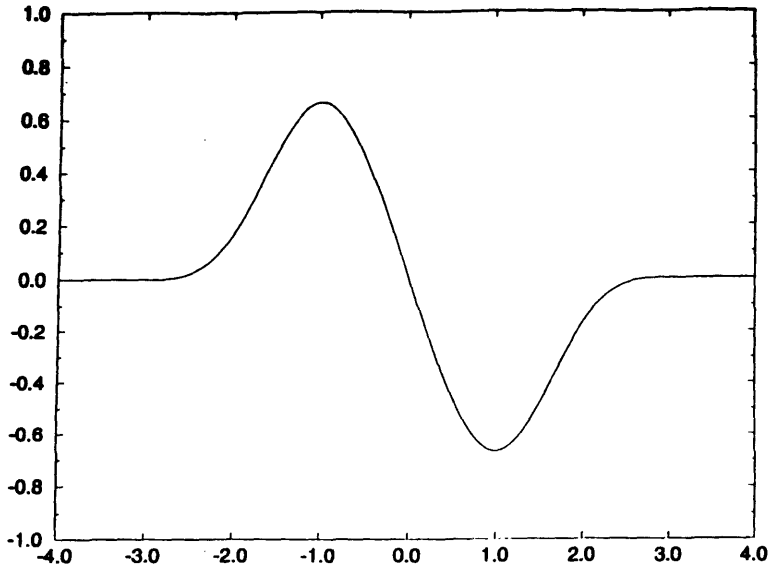


(a) Wavelet W1

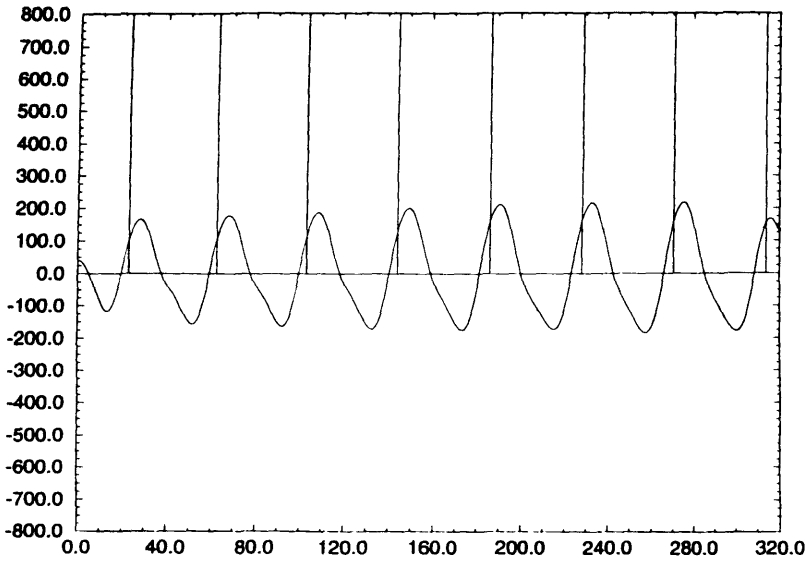


(b) Rescaled BSWT coefficients

Figure 5.4.3-1: Wavelet W1 and its rescaled BSWT coefficients for clean test data. (a) Wavelet W1, (b) Rescaled BSWT coefficients.

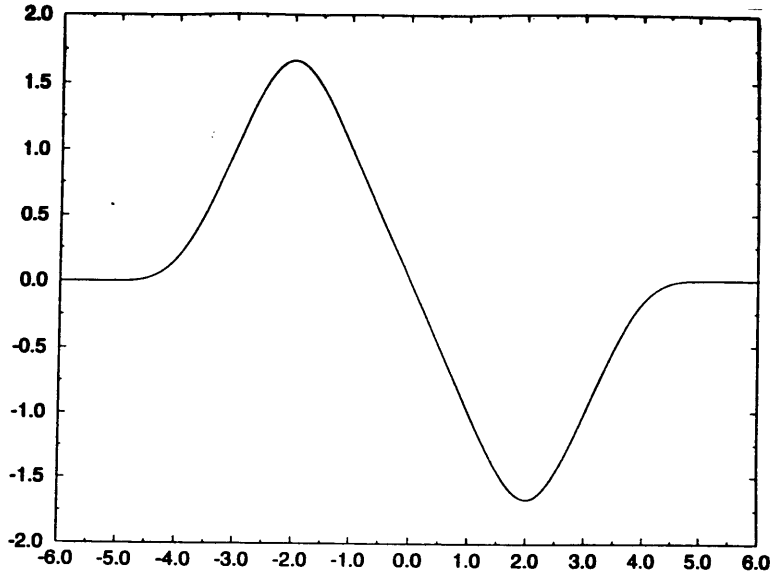


(a) Wavelet WA

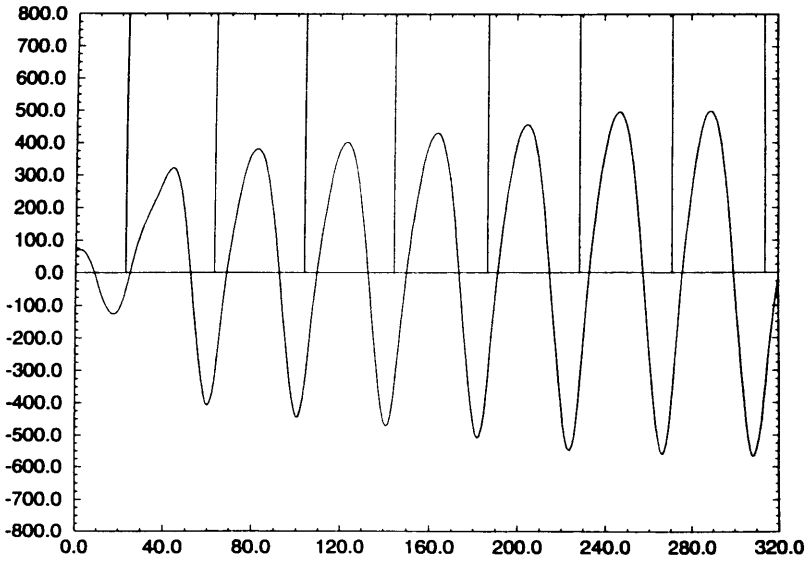


(b) BSWT coefficients

Figure 5.4.3-2: Wavelet WA and its rescaled BSWT coefficients for clean test data. (a) Wavelet WA, (b) BSWT coefficients.

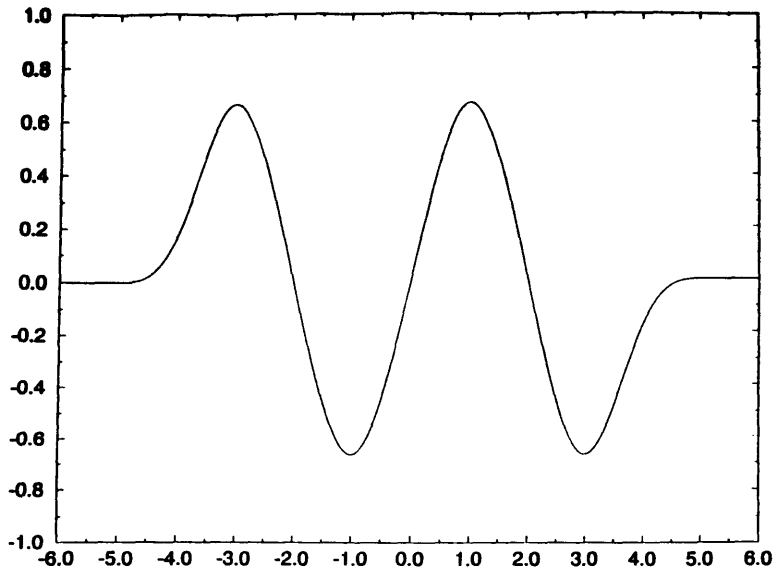


(a) Wavelet WB

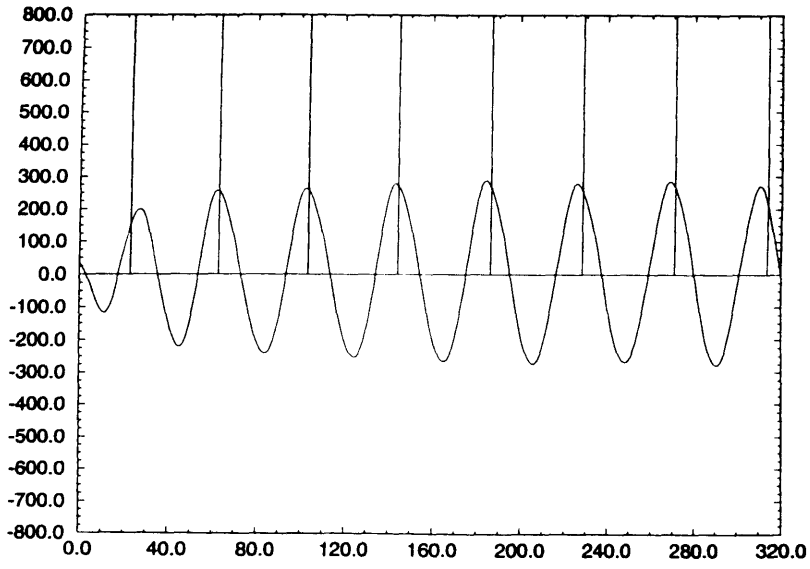


(b) BSWT coefficients

Figure 5.4.3-3: Wavelet WB and its rescaled BSWT coefficients for clean test data. (a) Wavelet WB, (b) BSWT coefficients.

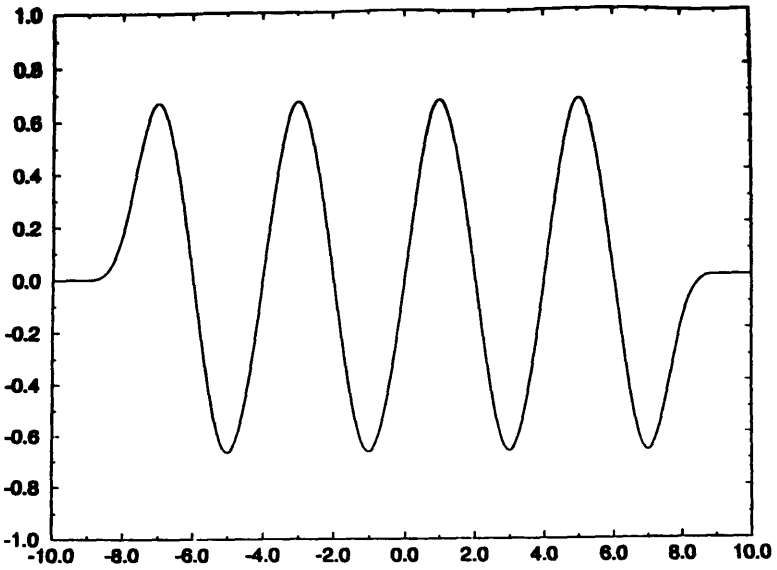


(a) Wavelet WC

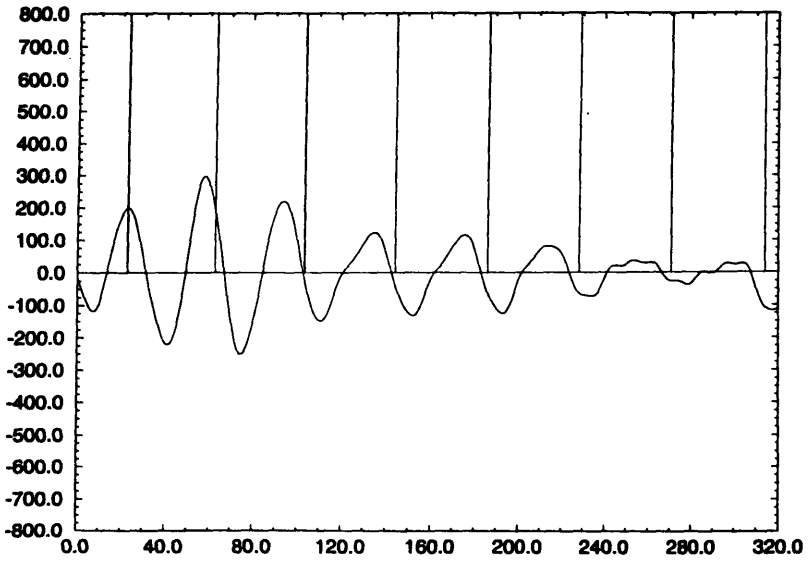


(b) BSWT coefficients

Figure 5.4.3-4: Wavelet WC and its rescaled BSWT coefficients for clean test data. (a) Wavelet WC, (b) BSWT coefficients.

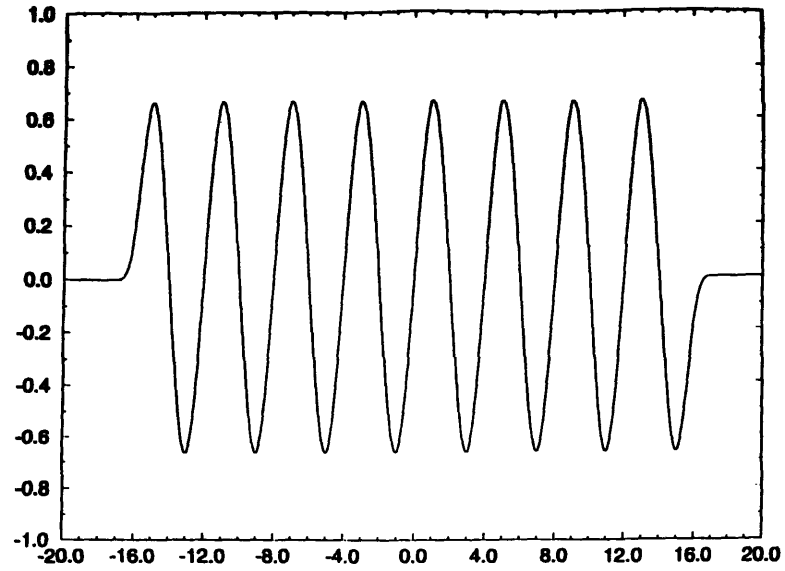


(a) Wavelet WD

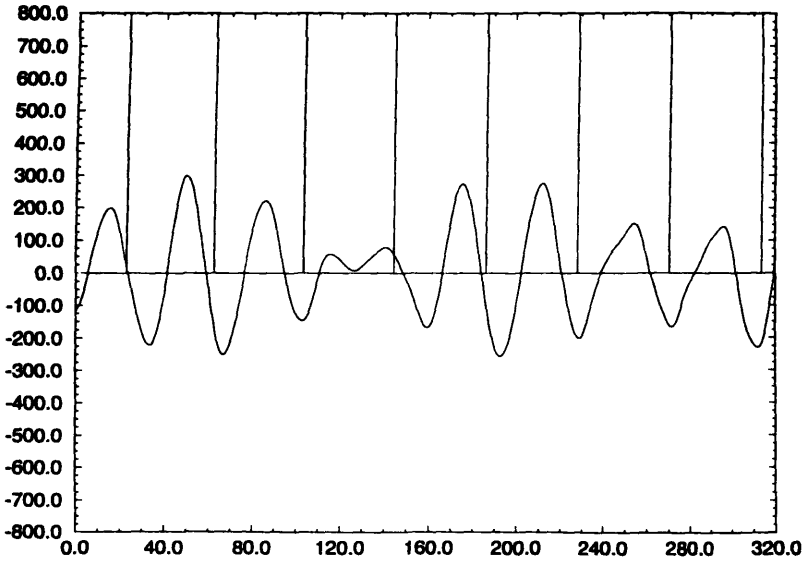


(b) BSWT coefficients

Figure 5.4.3-5: Wavelet WD and its rescaled BSWT coefficients for clean test data. (a) Wavelet WD, (b) BSWT coefficients.



(a) Wavelet WE



(b) BSWT coefficients

Figure 5.4.3-6: Wavelet WE and its rescaled BSWT coefficients for clean test data. (a) Wavelet WE, (b) BSWT coefficients.

5.4.4 Noisy Speech

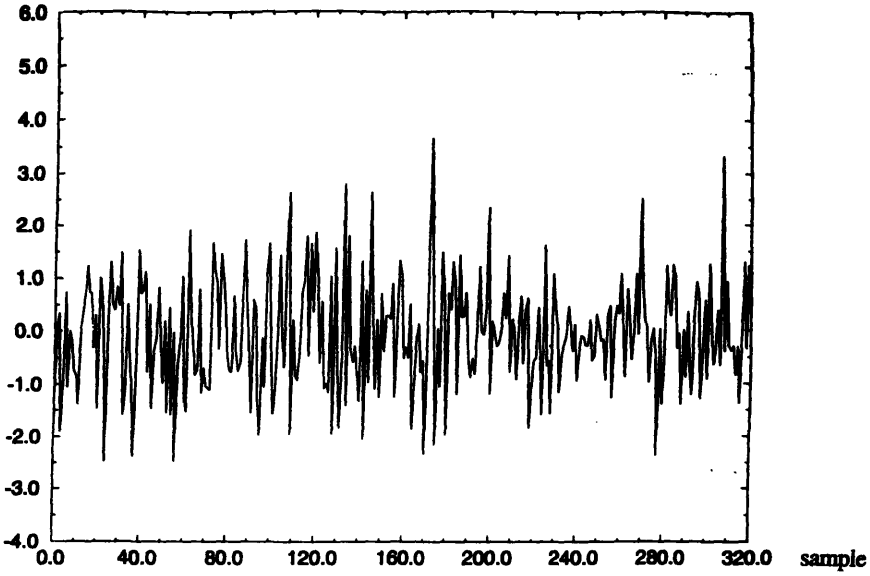
One of the most formidable problems in speech research is accurate pitch detection in a noisy environment. A few experiments are set up to see how this problem can be approached with our W1 wavelet.

The noisy test data are obtained by adding appropriately scaled white Gaussian noise to the clean speech samples (Fig. 5.2-3b). First, we need to verify that the lack of pitch structure (i.e., repetition in the waveform) of the Gaussian noise in the time domain is maintained in the transformed domain. Fig. 5.4.4-1 shows both the 320 samples of noise and the resulting BSWT coefficients based on W1. The lack of pitch structure is, indeed, preserved.

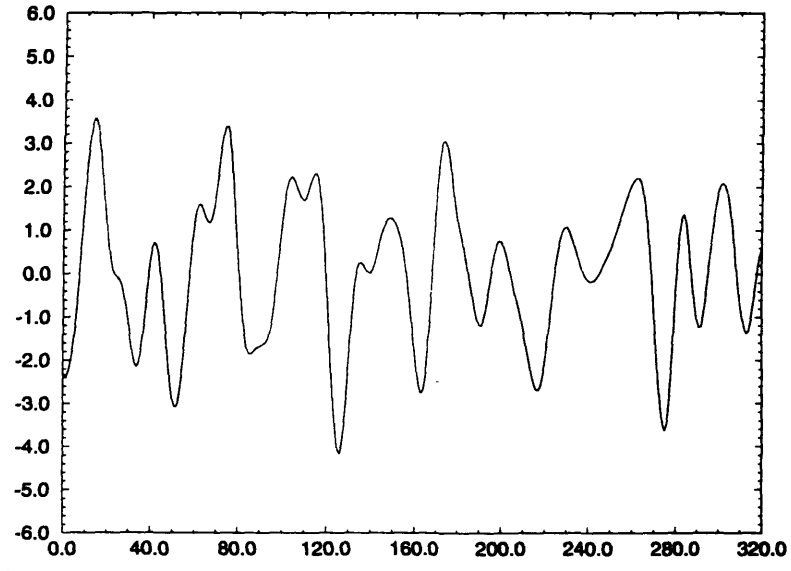
To study how W1 performs in noise, we evaluate the transform at eight different SNR test speech data (See Figs 5.2-4a to 5.2-4h). The results are shown in Figs. 5.4.4-2a to Fig. 5.4.4-2h. The noise has negligible effect until at 5 dB SNR when a spurious pitch peak is introduced, as illustrated by the lobe pointed to by the arrow. At 4 dB SNR, the lobe becomes more evident. To correctly determine the pitch period, a peak-picking algorithm needs to be more complex: it needs to have some logic to discount false local maxima. Let us now investigate ways to alleviate some of the complexity of the peak-picking algorithm.

When the noise level is high enough, its waveform looks very much like speech, in which case confusing pitch peaks must be differentiated from those due to genuine speech. The antisymmetric wavelet W1 has done a good job of detecting the actual pitch onset, but to suppress the effects of false pitch, the wavelet has to be redesigned. We propose to do this by destroying the antisymmetry of the wavelet. Two wavelets, Wα and Wβ, are designed; their *q[k]* parameters and shapes are shown in Fig. 5.4.4-3a and Fig. 5.4.4-4a,

respectively. The new wavelets are tested with SNR data of 5 dB and 4 dB; the results can be found in parts (b) and (c) of Figs. 5.4.4-3 and 5.4.4-4. We see that the new wavelets do provide some improvement because the lobe is sufficiently smoothed out so that there is no longer a spurious local maximum. In addition, because the actual pitch onset still causes the greatest excitation, its local maximum is still preserved. The cost, however, is a shift of the predicted pitch peaks away from the actual ones.



(a) 320 samples of Gaussian noise



(b) Wavelet W1 BSWT coefficients

Figure 5.4.4-1: Gaussian noise samples as test data and wavelet W1 BSWT coefficients.
(a) 320 samples of Gaussian noise.
(b) Wavelet W1 BSWT coefficients.

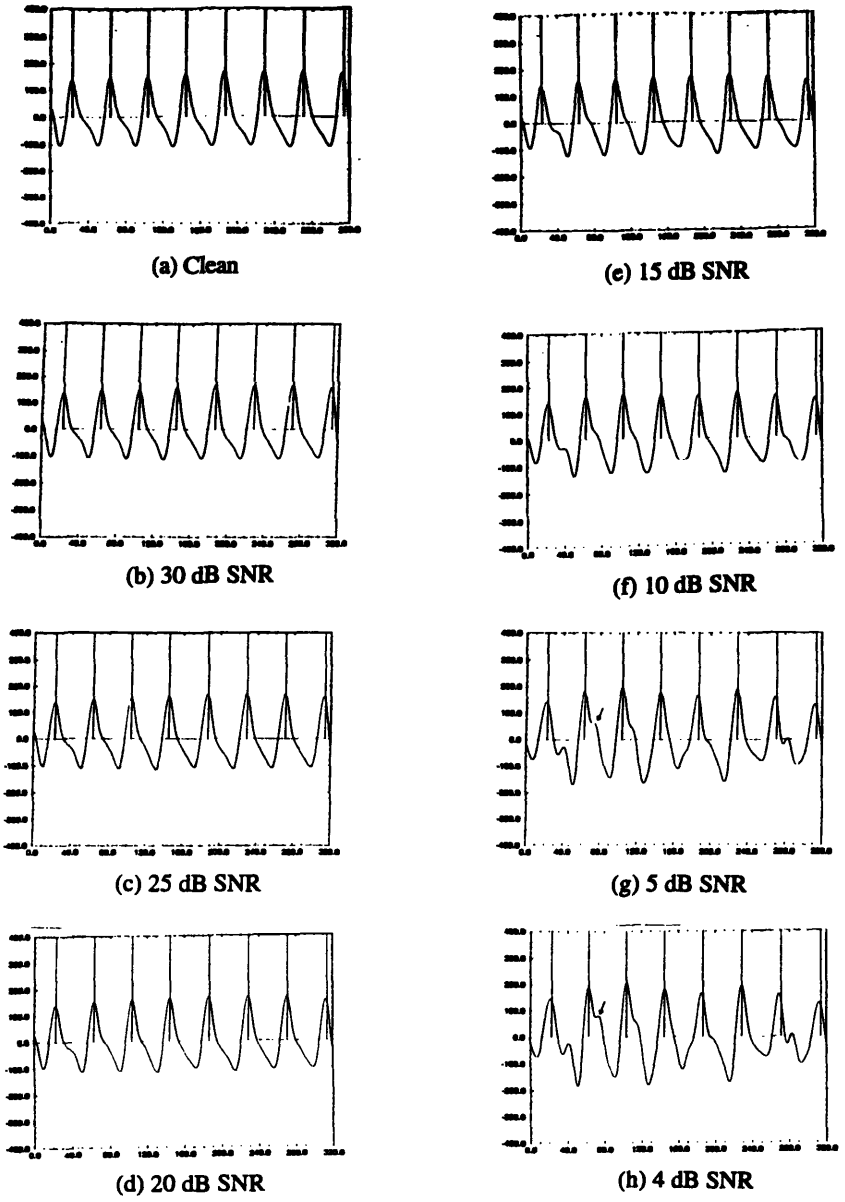
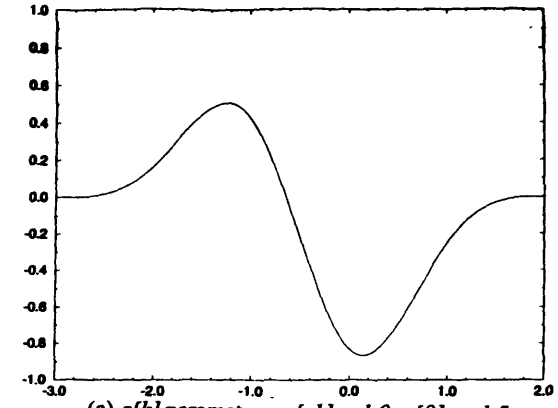
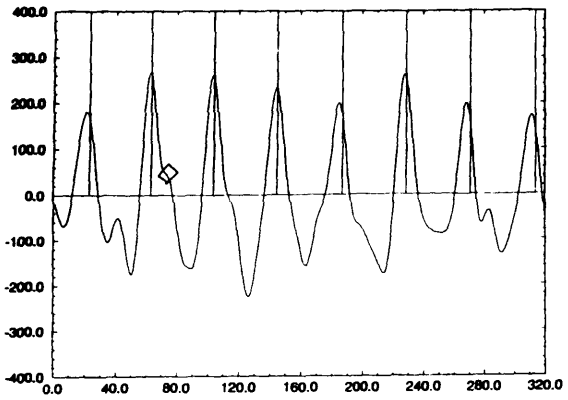


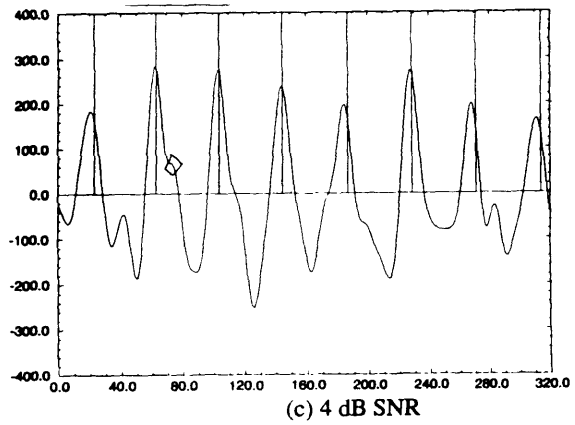
Figure 5.4.4-2: Wavelet W1 BSWT coefficients for eight different test data SNR's.
(a) Clean, (b) 30 dB SNR, (c) 25 dB SNR, (d) 20 dB SNR,
(e) 15 dB SNR, (f) 10 dB SNR, (g) 5 dB SNR, (h) 4 dB SNR.



(a) *q[k]* parameters: *q[-1] = 1.0, q[0] = -1.5*

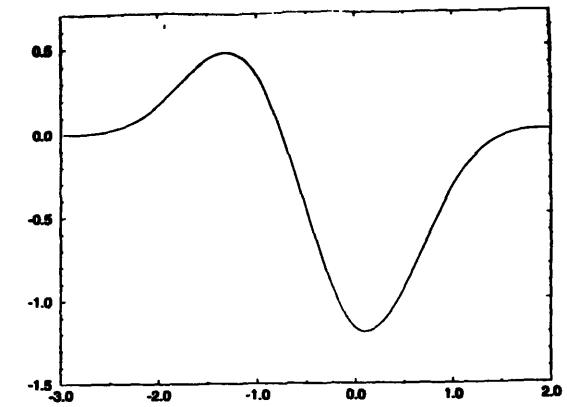


(b) 5 dB SNR

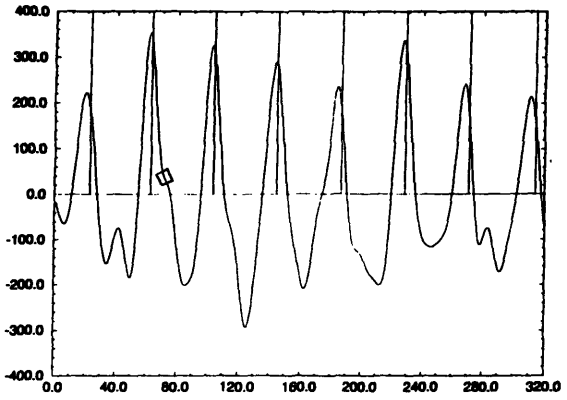


(c) 4 dB SNR

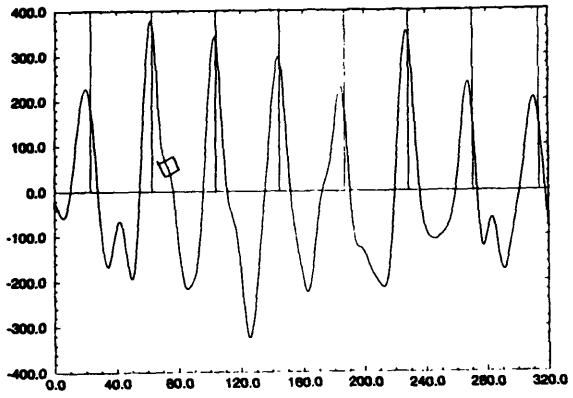
Figure 5.4.4-3: Wavelet W α and its performance in high noise test data.
(a) Wavelet W α *q[k]* parameters and shape,
(b) BSWT coefficients for 5 dB SNR test data,
(c) BSWT coefficients for 4 dB SNR test data.



(a) *q[k]* parameters: *q[-1] = 1.0, q[0] = -2.0*



(b) 5 dB SNR



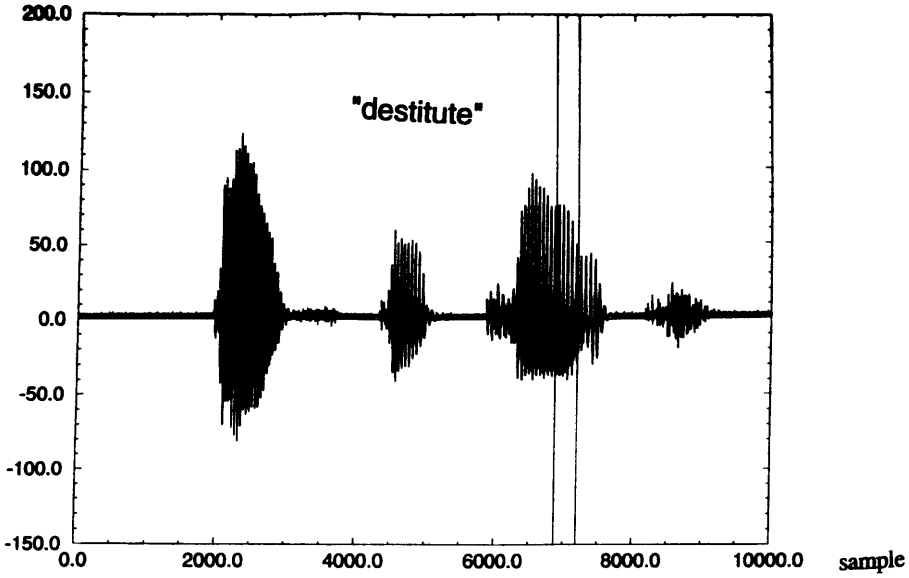
(c) 4 dB SNR

Figure 5.4.4-4: Wavelet Wβ and its performance in high noise test data.
(a) Wavelet Wβ *q[k]* parameters and shape,
(b) BSWT coefficients for 5 dB SNR test data,
(c) BSWT coefficients for 4 dB SNR test data.

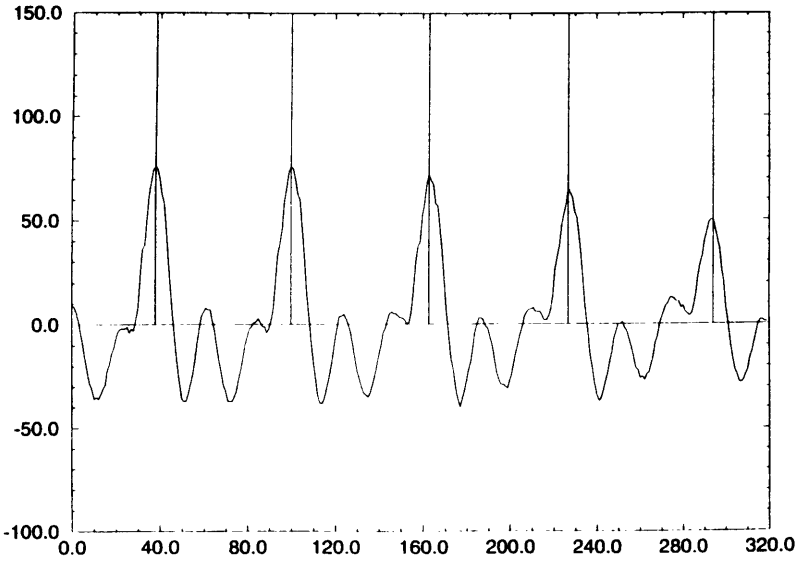
5.4.5 Multi-speaker speech

Pitch detection in multi-speaker speech is approached from the standpoint that all speech except the one from the main speaker are to be viewed as noise (See Section 5.4.4). In this section's experiments, the speech data are produced by two speakers, designated as main and background. Speech samples of the main speaker are shown in Fig. 5.2-3b; those of the background speaker are made up of 320 samples (Fig. 5.4.5-1b) between the two vertical lines of demarcation in the word "destitute" (Fig. 5.4.5-1a). Seven different SNR testing data are obtained by adding scaled background speech samples to the main speech samples (See Figs. 5.4.5-2a to 5.4.5-2g). The BSWT coefficients based on the wavelet W1 are evaluated for these test speech samples (See Figs. 5.4.5-3a to 5.4.5-3g). At 10 dB SNR and lower, pitch contributed by the background speech starts to appear. At 4 dB SNR, these background pitch peaks become very prominent.

Since wavelet $W\beta$ (Fig. 5.4.4-4a) exhibits better performance for noisy speech in the experiments of Section 5.4.4, let us reuse it here. The BSWT coefficients based on $W\beta$ for the three SNR test data of 10 dB, 5 dB, and 4 dB are evaluated and are shown in Fig. 5.4.5-4. Significant improvement is achieved for 10 dB SNR; for 5 dB and 4 dB SNR, improvement is very limited, though background pitch peaks very close to the actual ones are suppressed. In these cases, the peak-picking algorithm must include some kind of post pitch processing. For example, in cases when one predicted pitch period is very different from both of its neighbors, one of its neighbors' pitch period may be used to take its place.



(a) Word: "destitute"



(b) Test data of 320 samples

Figure 5.4.5-1: Background speech.
(a) Word: "destitute"
(b) Test data of 320 samples between the two vertical lines of "destitute"

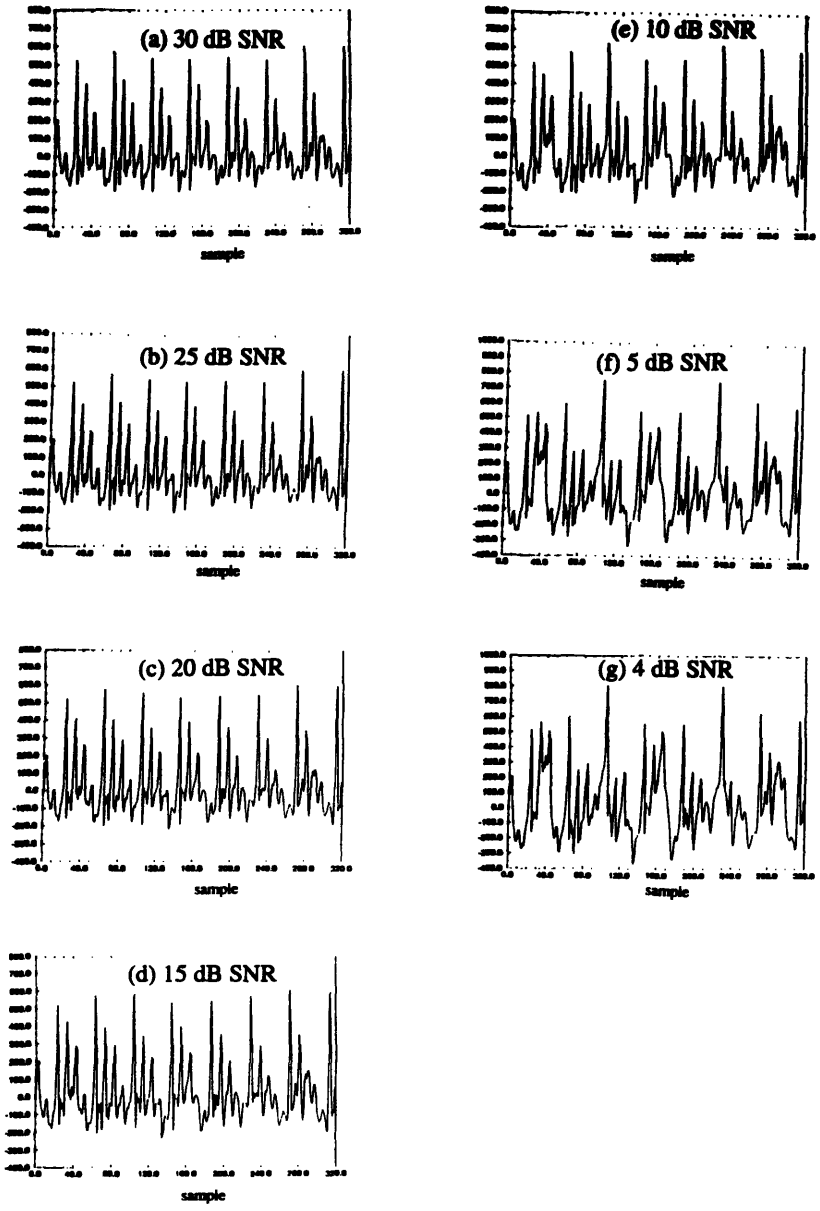
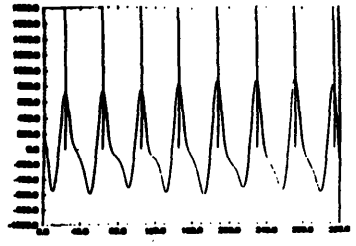
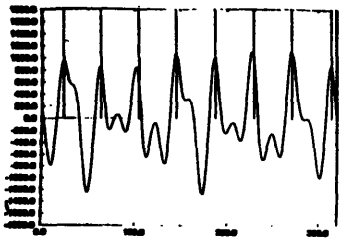


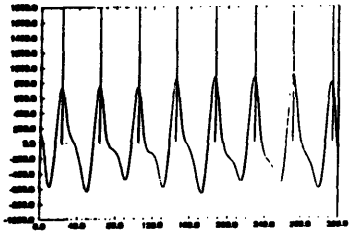
Figure 5.4.5-2: Dual-speaker speech test data at seven different SNR's. At
(a) 30 dB SNR, (b) 25 dB SNR, (c) 20 dB SNR
(d) 15 dB SNR, (e) 10 dB SNR, (f) 5 dB SNR, (g) 4 dB SNR.



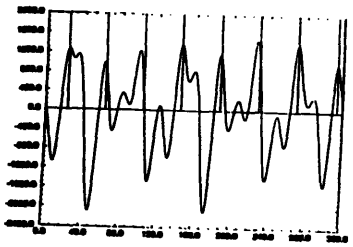
(a) 30 dB SNR



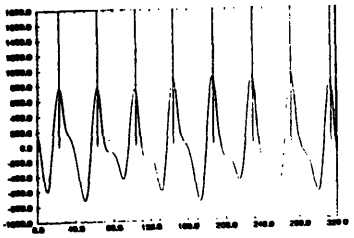
(e) 10 dB SNR



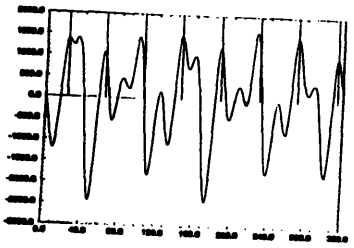
(b) 25 dB SNR



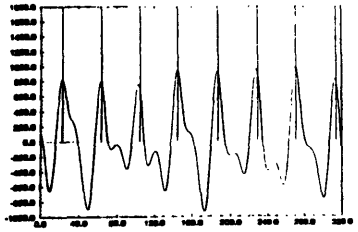
(f) 5 dB SNR



(c) 20 dB SNR

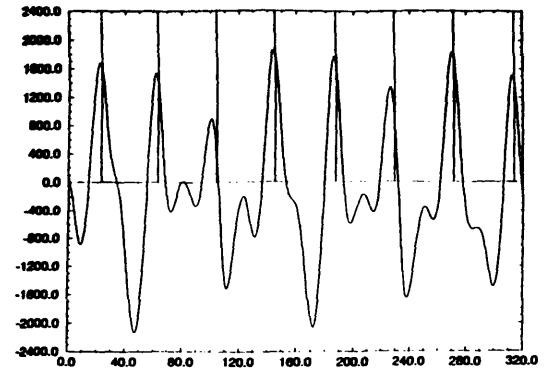


(g) 4 dB SNR

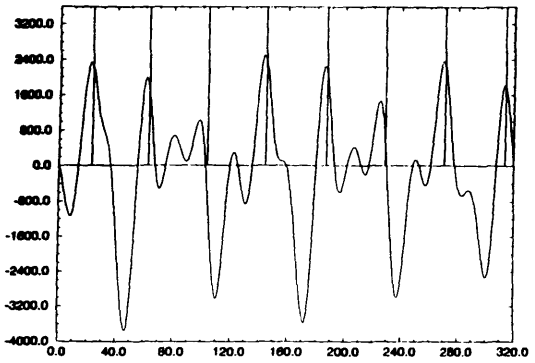


(d) 15 dB SNR

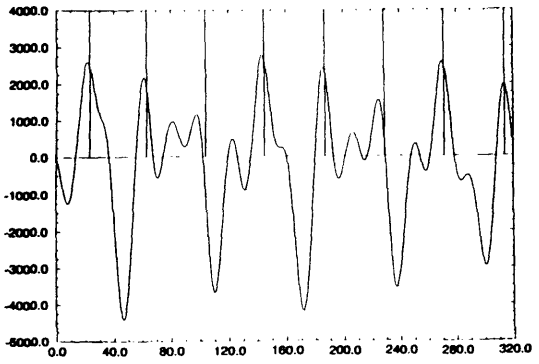
Figure 5.4.5-3: BSWT coefficients based on wavelet W1 for seven different SNR dual-speaker test data. For
(a) 30 dB SNR, (b) 25 dB SNR, (c) 20 dB SNR
(d) 15 dB SNR, (e) 10 dB SNR, (f) 5 dB SNR, (g) 4 dB SNR.



(a) 10 dB SNR



(b) 5 dB SNR



(c) 4 dB SNR

Figure 5.4.5-4: BSWT coefficients based on the wavelet $W\beta$ for dual-speaker test data. (Background speech is taken as noise.) For
(a) 10 dB SNR, (b) 5 dB SNR, (c) 4 dB SNR.

5.4.6 Unvoiced Speech

An interesting experiment to perform involves unvoiced speech, which, like Gaussian noise, has no observable pitch structure. We expect this property to be preserved in the transformed domain. Let us check if this is true.

Fig. 5.4.6-1a presents a discrete-time plot of the word "substance" and Fig. 5.4.6-1b shows the test data which consist of 320 samples in the ending unvoiced \s\ sound. The BSWT coefficients based on the W1 wavelet are plotted in Fig. 5.4.6-2. As expected, there is, indeed, no pitch structure.

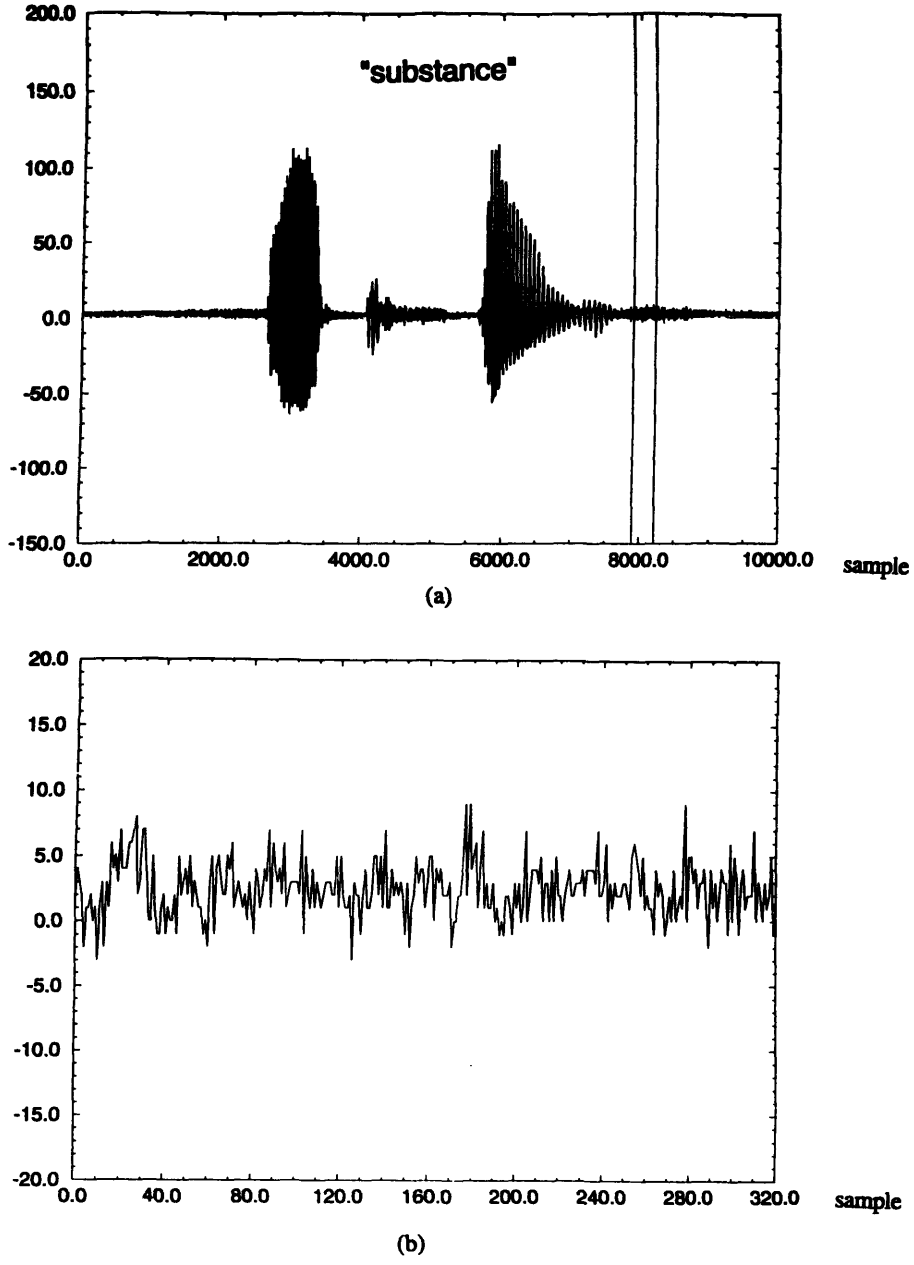


Figure 5.4.6-1: Unvoiced speech test data.
(a) Word: "substance"
(b) 320 samples of unvoiced speech samples between the two vertical lines in (a).

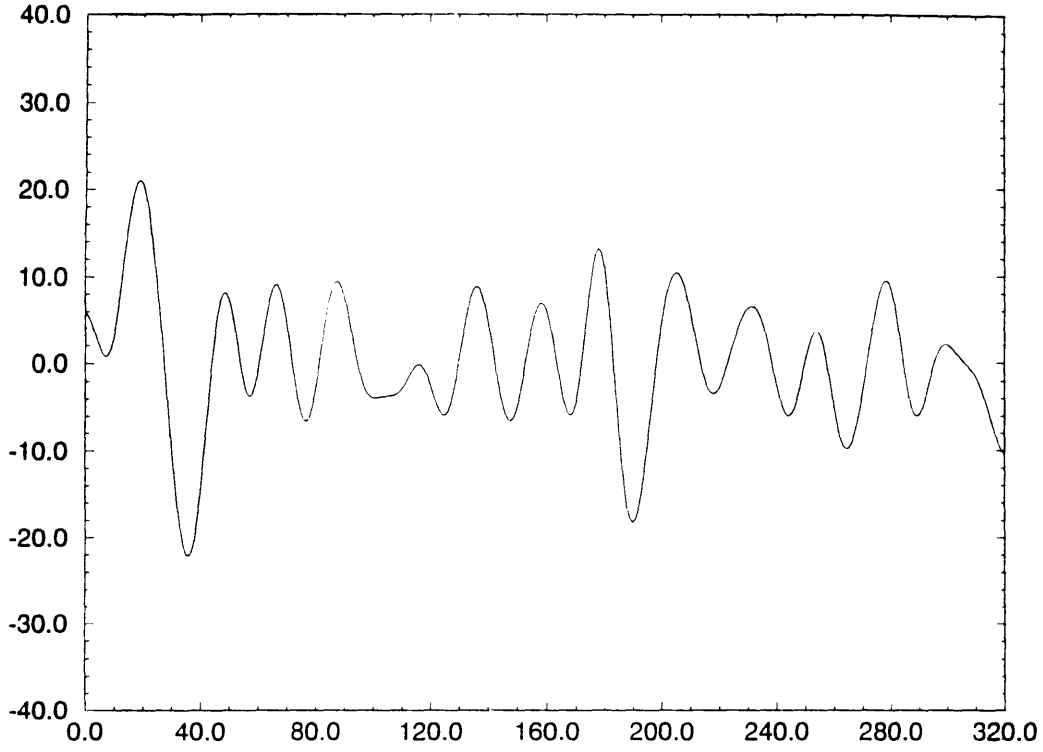


Figure 5.4.6-2: BSWT coefficients based on wavelet W1 for 320 samples of unvoiced test data.

Chapter 6 Conclusion

In previous chapters, we have investigated the wavelet transform, the spline functions, the transform based on B-spline functions, and various important factors that must be considered when the transform is employed in pitch detection. The wavelet transform is a very powerful tool, and from its definition alone, we see that it can be used to examine signals in detail by using various scales and time shifts. With a specific choice of a mother wavelet, we can magnify an interesting but hard-to-detect part of a signal (e.g., an edge of an image or the instant of glottal closure of a speech signal) so that further processing can be simplified. However, to achieve a meaningful and useful magnification, the "correct" mother wavelet must be chosen. Using wavelets built from B-splines, we have set out to determine what "correctness" of a wavelet entails in the application of pitch detection of a speech signal.

From the experimental results, a few conclusions can be drawn. First, the wavelet must be very flexible. Like a speech signal, it must be both positive and negative. And since we are detecting the instant of glottal closure, an event of very short duration, the wavelet should have minimal support. A good conversion of the non-stationary event of glottal closure into local maxima in the transform coefficients is obtained when the wavelet is approximated with cubic splines and the transform is evaluated at a scale of eight. We have found that a compactly supported, non-orthogonal, antisymmetric wavelet provides satisfactory results when there is no background interference. However, when the background is either noise or another speaker, we have found that a slight change in the wavelet's shape so that it is no longer antisymmetric can cause the true pitch pulses to be discriminated from false ones, at some cost in the accuracy of estimating pitch pulse times.

A major consequence of these experiments is a new strategy to approach the problem of pitch detection in adverse environments. Other consequences come from the plot of the transform coefficient maxima. Since these points of maxima coincide with the actual instants of glottal closure, we see that the transform has great potential in applications such as pitch-synchronous speech coding. In addition, the difficult problem of determining the quasi-periodicity in a speech frame is avoided because we are working with the exact location of pitch pulses in a frame, not the frame as a whole. In addition, from the results of the unvoiced test data, we see that the transform can also be used for voiced/unvoiced speech sound classification by taking advantage of the fact that no periodicity exists in a frame of unvoiced speech.

As for the consequence of this entire work, we have found that the wavelet transform is a very general tool, and in a specific application, there is much freedom in deciding how the transform can be used. We have illustrated the decision-making process for one application, and are hopeful that a similar process can be used for others.

Appendix A

Mathematical notations and definitions :

- $\Re$ denotes the set of real numbers, $\{-\infty, \infty\}$.
- $L^1(\Re)$ denotes the space of absolutely integrable functions defined on $\Re$, such that a function, $f(t)$, belongs to this space if
$$\int_{-\infty}^{\infty} |f(t)|\, dt \quad < \quad \infty.$$
- $L^2(\Re)$ denotes Hilbert space, or the space of square-integrable functions defined on $\Re$, such that a function, $f(t)$, belongs to this space if
$$\int_{-\infty}^{\infty} |f(t)|^2\, dt \quad < \quad \infty$$
- $H^2(\Re)$, denotes Hardy space such that a function, $f(t)$, belongs to this space if
$$(1) \quad f(t) \quad \in \quad L^2(\Re),$$
$$(2) \quad \hat{f}(\omega) \quad = \quad 0 \quad for \quad \omega < 0.$$
- $L^2(0, T)$ denotes the space of T-periodic square-integrable functions defined on $(0, T)$, such that a function, $f(t)$, belongs to this space if
$$\int_{0}^{T} |f(t)|^2 dt \quad < \quad \infty.$$
- *Inner product* of two functions $f(t)$ and $g(t)$ is $\langle f, g \rangle = \int_{-\infty}^{\infty} f(t)\, \overline{g(t)}\, dt$; in $L^2(\Re)$, the product simply becomes $\langle f, g \rangle = \int_{-\infty}^{\infty} f(t)\, g(t)\, dt$.
- *Norm* of $L^2(\Re)$: $\|f\|_2 \equiv \langle f, f \rangle^{\frac{1}{2}} = \sqrt{\int_{-\infty}^{\infty} |f(t)|^2\, dt}$
- *Support* of function $f(t)$ is defined as $\{t \in \Re: f(t) \neq 0\}$.

Compact support denotes a support that is relatively small.

- *Characteristic function* of interval $[t_1, t_2]$ is $\chi_{t_1,t_2}(t) = \begin{cases} 1, & t \in [t_1, t_2] \\ 0, & otherwise \end{cases}$
- *Kronecker delta* is $\delta_{xy} = \begin{cases} 1, & x = y \\ 0, & otherwise \end{cases}$
- Given a space of $L^2(\Re)$ with inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$. If $\{w_n\}$ denotes a sequence of elements of the space, then

  Span of $\{w_n\} = \{ \sum_{n_1}^{n_2} c_n w_n : n_1 < 0, \ n_2 > 0, \ c_n \in$ *set of complex numbers* $\}$.

  - If $\langle w_{n_1}, w_{n_2} \rangle = 0$ and $n_1 \neq n_2$, then $\{w_n\}$ is *orthogonal.*
  - If $\{w_n\}$ is orthogonal and $\|w_n\| = 1$, then $\{w_n\}$ is *orthonormal.*
  - If $w = 0$ is the only element in $L^2(\Re)$ that is orthogonal to $\{w_n\}$, $\{w_n\}$ is *complete* in $L^2(\Re)$.

- Z denotes the set of integers, $\ldots, -2, -1, 0, 1, 2, \ldots$
- $\ell^1(Z)$ denotes the space of absolutely summable sequences such that a sequence, $c(n)$, belongs to this space if

$$\sum_{n=-\infty}^{\infty} |c(n)| < \infty.$$

- $\ell^2(Z)$ denotes the space of square-summable sequences such that a sequence, $c(n)$, belongs to this space if

$$\sum_{n=-\infty}^{\infty} |c(n)|^2 < \infty.$$

- *Regularity* and *smoothness* of a function are a measure of the number of continuous derivatives it possesses.

Appendix B

Let's see how the admissibility condition can be derived from the following relation

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left|\left\langle e^{u/2}\,\psi(e^{u}t-v),\quad \psi(t)\right\rangle\right|^{2} du\,dv \quad < \quad \infty. \tag{B-1}$$

Since the Fourier transform is an unique one-to-one mapping, taking the FT's of the two functions inside the inner product of (B-1), we obtain

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left|\left\langle e^{-u/2}\,e^{-iv\omega}\,\hat{\psi}(e^{-u}\omega),\quad \hat{\psi}(\omega)\right\rangle\right|^{2} du\,dv \quad < \quad \infty. \tag{B-2}$$

Expanding first the inner product and then the squared term in the integrand results in

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left|\int_{-\infty}^{\infty} e^{-u/2}\,e^{iv\omega}\,\overline{\hat{\psi}(e^{-u}\omega)}\,\hat{\psi}(\omega)\,d\omega\right|\left|\int_{-\infty}^{\infty} e^{-u/2}\,e^{-iv\omega}\,\hat{\psi}(e^{-u}\omega)\,\overline{\hat{\psi}(\omega)}\,d\omega\right| du\,dv \quad < \quad \infty. \tag{B-3}$$

Substituting the inverse FT relation,

$$\psi(v) \quad = \quad \frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{\psi}(\omega)\,e^{iv\omega}\,d\omega \tag{B-4}$$

into (B-3) yields

$$2\pi\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left|e^{-u/2}\,\psi(v)\,\overline{\hat{\psi}(e^{-u}\omega)}\right|\left|\int_{-\infty}^{\infty} e^{-u/2}\,e^{-iv\omega}\,\hat{\psi}(e^{-u}\omega)\,\overline{\hat{\psi}(\omega)}\,d\omega\right| du\,dv \quad < \quad \infty. \tag{B-5}$$

Replacing $\hat{\psi}(\omega)$ in the second absolute expression in the integrand of (B-5) with

$$\hat{\psi}(\omega) \quad = \quad \int_{-\infty}^{\infty}\psi(v)\,e^{-iv\omega}\,dv \tag{B-6}$$

renders

$$2\pi\int_{-\infty}^{\infty} e^{-u}\left|\hat{\psi}(e^{-u}\omega)\right|^{2} du\int_{-\infty}^{\infty}\left|\hat{\psi}(\omega)\right|^{2} d\omega \quad < \quad \infty. \tag{B-7}$$

Substituting $x = e^{-u}\omega$ into the first integral, we have

$$2\pi\int_{-\infty}^{\infty}\left|\hat{\psi}(x)\right|^{2} dx\int_{-\infty}^{\infty}\frac{\left|\hat{\psi}(\omega)\right|^{2}}{\omega}d\omega \quad < \quad \infty. \tag{B-8}$$

Because $\psi(t) \in H^{2}(\Re)$, with Parseval's relation we can see that the first integral is finite. Hence, the entire expression is less than infinity as long as

$$\int_0^{\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega} d\omega \quad < \quad \infty. \tag{B-9}$$

For a function $\psi(t) \in L^2(\Re)$, the condition in (B-9) translates into the admissibility condition.

Appendix C

The following is a proof of the energy preservation equation of the CWT such that for $f(t) \in L^2(\Re)$ and $\psi(t)$ is admissible,

$$\frac{1}{c_\psi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|(CWT_t f)(b,a)|^2 \frac{1}{a^2}\,da\,db \;=\; \int_{-\infty}^{\infty}|f(t)|^2\,dt, \tag{C-1}$$

where

$$c_\psi \;=\; \int_{-\infty}^{\infty}\frac{|\hat{\psi}(\omega)|^2}{|\omega|}\,d\omega. \tag{C-2}$$

To show (C-1), let us carry out the two integrals on the left-hand side of (C-1) separately, first with respect to the differential *db*, then with respect to *da*, and finally, check if the two sides match.

Expanding the integrand on the left-hand side of (C-1),

$$\int_{-\infty}^{\infty}|(CWT_t f)(b,a)|^2\,db \;=\; \int_{-\infty}^{\infty}[((CWT_t f)(b,a))\;\overline{((CWT_t f)(b,a))}]db. \tag{C-3}$$

Rewriting the two distinct CWT integrals on the right-hand side of (C-3) explicitly, we have

$$\frac{1}{|a|}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(t)\,\overline{\psi(\frac{t-b}{a})}\,dt\;\int_{-\infty}^{\infty}\overline{f(r)}\,\psi(\frac{r-b}{a})\,dr\;db. \tag{C-4}$$

Substituting

$$f(t) \;=\; \frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{f}(\omega)e^{i\omega t}\,d\omega$$

and

$$\overline{f(r)} \;=\; \frac{1}{2\pi}\overline{\int_{-\infty}^{\infty}\hat{f}(\omega)e^{i\omega r}\,d\omega}$$

into (C-4) yields

$$\frac{1}{|a|}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{f}(\omega)e^{i\omega t}\,d\omega\;\overline{\psi(\frac{t-b}{a})}\,dt\;\int_{-\infty}^{\infty}\frac{1}{2\pi}\overline{\int_{-\infty}^{\infty}\hat{f}(\omega)e^{i\omega r}\,d\omega}\;\psi(\frac{r-b}{a})\,dr\,db. \tag{C-5}$$

Using

$$\int_{-\infty}^{\infty} \psi(\frac{t-b}{a}) e^{-i\omega t}\, dt \;=\; a e^{-ib\omega}\, \hat{\psi}(a\omega), \tag{C-6}$$

(C-5) can be simplified as

$$\frac{1}{|a|}\int_{-\infty}^{\infty} \frac{1}{2\pi}\overline{\int_{-\infty}^{\infty} \hat{f}(\omega)d\omega}\ \int_{-\infty}^{\infty} \psi(\frac{t-b}{a}) e^{-i\omega t}\, dt\ \ \frac{1}{2\pi}\int_{-\infty}^{\infty} \overline{\hat{f}(\omega)}d\omega\ \int_{-\infty}^{\infty} \psi(\frac{r-b}{a}) e^{-i\omega r}\, dr\ db \tag{C-7}$$

$$= \frac{1}{|a|}\int_{-\infty}^{\infty} \frac{1}{2\pi}\overline{\int_{-\infty}^{\infty} \hat{f}(\omega)d\omega a e^{-i\omega b}\, \hat{\psi}(a\omega)}\ \ \frac{1}{2\pi}\int_{-\infty}^{\infty} \overline{\hat{f}(\omega)}d\omega a e^{-i\omega b}\, \hat{\psi}(a\omega)\ db \tag{C-8}$$

$$= \frac{1}{|a|}\int_{-\infty}^{\infty} \frac{a}{2\pi}\overline{\int_{-\infty}^{\infty} \hat{f}(\omega)\, \hat{\psi}(a\omega) e^{-i\omega b}\, d\omega}\ \ \frac{a}{2\pi}\int_{-\infty}^{\infty} \overline{\hat{f}(\omega)}\, \hat{\psi}(a\omega) e^{-i\omega b}\, d\omega\ db. \tag{C-9}$$

Now let $P(\omega) = \overline{\hat{f}(\omega)}\, \hat{\psi}(a\omega)$, then (C-9) can be rewritten as

$$\frac{1}{|a|}\int_{-\infty}^{\infty} \frac{a}{2\pi}\overline{\int_{-\infty}^{\infty} P(\omega) e^{-i\omega b}\, d\omega}\ \ \frac{a}{2\pi}\int_{-\infty}^{\infty} P(\omega) e^{-i\omega b}\, d\omega\ db \tag{C-10}$$

$$= \frac{1}{|a|}\int_{-\infty}^{\infty} \frac{a}{2\pi}\overline{\hat{P}_b(\omega)}\ \ \frac{a}{2\pi}\hat{P}_b(\omega)\ db, \tag{C-11}$$

where

$$\hat{P}_b(\omega) \;=\; \int_{-\infty}^{\infty} P(\omega) e^{-i\omega b}\, d\omega. \tag{C-12}$$

After an application of Parseval's Relation, (C-11) can be simplified as

$$\frac{a^2}{|a|}\frac{1}{2\pi}\int_{-\infty}^{\infty} \overline{P(\omega)} P(\omega) d\omega \tag{C-13}$$

$$= \frac{a^2}{2\pi|a|}\int_{-\infty}^{\infty} \left|\hat{f}(\omega)\right|^2 \left|\hat{\psi}(a\omega)\right|^2 d\omega. \tag{C-14}$$

Now if we integrate with respect to the differential $\frac{1}{a^2}da$, we have

$$\frac{1}{2\pi|a|}\int_{-\infty-\infty}^{\infty}\int^{\infty} \left|\hat{f}(\omega)\right|^2 \left|\hat{\psi}(a\omega)\right|^2 d\omega\, da \tag{C-15}$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} \left|\hat{f}(\omega)\right|^2 d\omega \int_{-\infty}^{\infty} \frac{\left|\hat{\psi}(a\omega)\right|^2}{|a|} da. \tag{C-16}$$

Let $s = a\omega$, then (C-16) can be rewritten as

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \left|\hat{f}(\omega)\right|^2 d\omega \int_{-\infty}^{\infty} \frac{\left|\hat{\psi}(s)\right|^2}{|s|} ds \tag{C-17}$$

$$= c_\psi \frac{1}{2\pi}\left\langle \hat{f}(\omega), \hat{f}(\omega) \right\rangle, \tag{C-18}$$

where c_ψ is defined in (C-2). Invoking Parseval's Relation one more time results in the relation we set out to show, equation (C-1).

References

[1] Allen, J.B., "Short Term Spectral Analysis, Synthesis, and Modification by Discrete Fourier Transform," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-25, No. 3, June 1977, pp. 235-238.

[2] Allen, J.B., Rabiner, L.R., "A Unified Approach to Short-Time Fourier Analysis and Synthesis," *Proceedings of the IEEE*, Vol. 65, No. 11, November 1977, pp. 1558-1564.

[3] Schafer, R.W., Rabiner, L.R., "Design and Simulation of a Speech Analysis-Synthesis System Based on Short-Time Fourier Analysis," *IEEE Transactions on Audio and Electroacoustics*, Vol. AU-21, No. 3, June 1973, pp. 165-174.

[4] Portnoff, M.R., "Time-Frequency Representation of Digital Signals and Systems Based on Short-Time Fourier Analysis," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-28, No. 1, February 1980, pp. 55-69.

[5] Portnoff, M.R., "Short-Time Fourier Analysis of Sampled Speech," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-29, No.3, June 1981, pp. 364-373.

[6] Rioul, O., Vetterli, M., "Wavelets and Signal Processing," *IEEE Signal Processing Magazine*, October 1991, pp. 14-38.

[7] Hlawatsch, F., Boudreaux-Bartels, G.F., "Linear and Quadratic Time-Frequency Signal Representations," *IEEE Signal Processing Magazine*, April 1992, pp. 21-67.

[8] Cohen L., "Time-Frequency Distributions - A Review," Proceedings of the IEEE, Vol. 77, No. 7, July 1989, pp. 941-981.

[9] Friedlander, B., Porat, B., "Detection of Transient Signals by the Gabor Representation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 37, No. 2, February 1989, pp. 169-180.

[10] Wexler, J., Raz, S., "Discrete Gabor Expansions," *Signal Processing*, 21, 1990, pp. 207-220.

[11] Schafer, R.W., Rabiner, L.R., "Digital Representations of Speech Signals," *Proceedings of the IEEE*, Vol. 63, No. 4, April 1975, pp. 662-677.

[12] Chui, C.K., *An Introduction to Wavelets*, Academic Press, Boston, MA, 1992.

[13] Chui, C.K., (ed.), *Wavelets : A Tutorial in Theory and Applications*, Academic Press, Boston, MA, 1992.

[14] Oppenheim, A.V., Schafer, R.W., *Discrete-Time Signal Processing*, Prentice Hall, Englewood Cliffs, New Jersey, 1989.

[15] Rabiner, L.R., Schafer, R.W., *Digital Processing of Speech Signals*, Prentice-Hall, Englewood Cliffs, New Jersey, 1978.

[16] Cooke, M., Beet, S., Crawford, M., (ed.), *Visual Representation of Speech Signals*, John Wiley & Sons, Chichester, England, 1993.

[17] Sneddon, I.N., *Fourier Transforms*, McGraw-Hill Book Company, 1951.

[18] Zygmund, A., *Trigonometric Series*, vol. I and II, Cambridge University Press, 1968.

[19] Oppenheim, A.V., Willsky, A.S., *Signals and Systems*, Prentice-Hall, Endglewood Cliffs, New Jersey, 1983.

[20] Papoulis, A., *The Fourier Integral and its Applications*, McGraw-Hill Book Company, 1962.

[21] Boashash, B., *Time-Frequency Signal Analysis : Methods and Applications*, Longman Cheshire, Melbourne, Australia, 1992.

[22] Burrus, C.S., Gopinath, R.A., "Introduction to Wavelets and Wavelet Transforms," Rice University, Tutorial #1 Notes, April 26, 1993, ICASSP 1993 Minneapolis, MN.

[23] Daubechies, I., "Orthonormal Bases of Compactly Supported Wavelets," *Communications on Pure and Applied Mathematics*, Vol. XLI, 1988, pp. 909-996.

[24] Mallat, S.G., "Multiresolution Approximations and Wavelet Orthonormal Bases of $L^2(\Re)$," *Transactions of the American Mathematical Society*, Vol. 315, No. 1, September 1989, pp. 69-87.

[25] Daubechies, I., Grossman, A., Meyer, Y., "Painless nonorthogonal expansions," *J. Math. Phys.*, Vol. 27, No. 5, May 1986, pp. 1271-1283.

[26] Grossman, A., Morlet, J., "Decompositions of Hardy Functions into Square Integrable Wavelets of Constant Shape," *SIAM J. Math. Anal.*, Vol. 15, No. 4, July 1984, pp. 723-736.

[27] Mallat, S.G., "A Theory for Multiresolution Signal Decomposition : The Wavelet Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 7, July 1989, pp. 674-693.

[28] Strang, G., "Wavelets and Dilation Equations : A Brief Introduction," *SIAM Review*, Vol. 31, No. 4, December 1989, pp. 614-627.

[29] Heil, C.E., Walnut, D.F., "Continuous and Discrete Wavelet Transforms," *SIAM Review*, Vol. 31, No. 4, December 1989, pp. 628-666.

[30] Smith, M.J.T., Barnwell, T.P., "Exact Restruction Techniques for Tree-Structured Subband Coders," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-34, No. 3, June 1986, pp. 434-441.

[31] Krasner, M.A., "The Critical Band Coder -- Digital Encoding of Speech Signals Based on the Perceptual Requirements of the Auditory System," in *Proc. 1980 IEEE ICASSP*, Denver, CO, April 1980, pp. 327-331.

[32] Tuteur, F.B., "Wavelet Tranformations in Signal Detection," in *Proc. 1988 IEEE ICASSP*, New York, NY, D2.10, pp. 1435-1438.

[33] Mallat, S.G., "Multifrequency Channel Decompositions of Images and Wavelet Models," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 37, No. 12, December 1989, pp. 2091-2110.

[34] Daubechies, I., "The Wavelet Transform, Time-Frequency Localization and Signal Analysis," *IEEE Transactions on Information Theory*, Vol. 36, No. 5, September 1990, pp. 961-1005.

[35] Mallat, S., Zhong, S., "Characterization of Signals from Multiscale Edges," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, No. 7, July 1992, pp. 710-732.

[36] Vetterli, M., Herley, C., "Wavelets and Filter Banks: Theory and Design," *IEEE Transactions on Signal Processing*, Vol. 40, No. 9, September 1992, pp. 2207-2232.

[37] Vetterli, M., Herley, C., "Wavelets and Filter Banks: Relationships and New Results," in *Proc. 1990 IEEE ICASSP*, D12.1, pp. 1723-1726.

[38] Herley, C., Vetterli, M., "Linear Phase Wavelets: Theory and Design," in *Proc. 1991 IEEE ICASSP*, D9.1, pp. 2017-2020.

[39] Antonini, M., Barlaud, M., Mathieu, P., Daubechies, I., "Image Coding Using Vectore Quantization in the Wavelet Transform Domain," in *Proc. 1990 IEEE ICASSP*, M9.10, pp. 2297-2300.

[40] Herley, C., Vetterli, M., "Wavelets and Recursive Filter Banks," *IEEE Transactions on Signal Processing*, Vol. 41, No. 8, August 1993, pp. 2536-2556.

[41] Vaidyanathan, P.P., "Multirate Digital Filters, Filter Banks, Polyphase Networks, and Applications: A Tutorial," *Proceedings of the IEEE*, Vol. 78, No. 1, January 1990, pp. 56-93.

[42] DeVore, R.A., Jawerth, B., Lucier, B.J., "Image Compression Through Wavelet Transform Coding," *IEEE Transactions on Information Theory*, Vol. 38, No. 2, March 1992, pp. 719-746.

[43] Antoine, J.-P., Carrette, P., Murenzi, R., Piette, B., "Image analysis with two-dimensional continuous wavelet transform," *Signal Processing*, Vol. 31, No. 3, April 1993, pp. 241-272.

[44] Antonini, M., Barlaud, M., Mathieu, P., Daubechies, I., "Image Coding Using Wavelet Transform," *IEEE Transactions on Image Processing*, Vol. 1, No. 2, April 1992, pp. 205-220.

[45] Hou, H.S., Andrews, H.C., "Cubic Splines for Image Interpolation and Digital Filtering," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-26, No. 6, December 1978, pp. 508-517.

[46] Marsden, M.J., Richards, F.B., Riemenschneider, S.D., "Cardinal Spline Interpolation Operators on ℓ^p Data," *Indiana University Mathematics Journal*, Vol. 24, No. 7, 1975, pp. 677-689.

[47] Liou, M.-L., "Spline Fit Made Easy," *IEEE Transactions on Computers*, Vol. C-25, No. 5, May 1976, pp. 522-527.

[48] Horowitz, L.L., "The Effects of Spline Interpolation on Power Spectral Density," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-22, No. 1, February 1974, pp. 22-27.

[49] Unser, M., Aldroubi, A., Eden, M., "Fast B-Spline Transforms for Continuous Image Representation and Interpolation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, No. 3, March 1991, pp. 277-285.

[50] Unser, M., Aldroubi, A., Eden, M., "Polynomial Spline Signal Approximations: Filter Design and Asymptotic Equivalence with Shannon's Sampling Theorem," *IEEE Transactions on Information Theory*, Vol. 38, No. 1, January 1992, pp. 95-103.

[51] Unser, M., Aldroubi, A., Eden, M., "On the Asymptotic Convergence of B-Spline Wavelets to Gabor Functions," *IEEE Transactions on Information Theory*, Vol. 38, No. 2, March 1992, pp. 864-872.

[52] Aldroubi, A., Unser, M., Eden, M., "Cardinal spline filters: Stability and convergence to the ideal sinc interpolator," *Signal Processing*, Vol. 28, No. 2, August 1992, pp. 127-138.

[53] Unser, M., Aldroubi, A., Eden, M., "B-Spline Signal Processing: Part I--Theory," *IEEE Transactions on Signal Processing*, Vol. 41, No. 2, February 1993, pp. 821-833.

[54] Unser, M., Aldroubi, A., Eden, M., "B-Spline Signal Processing: Part II--Efficient Design and Applications," *IEEE Transactions on Signal Processing*, Vol. 41, No. 2, Frebruary 1993, pp. 834-848.

[55] Unser, M., Aldroubi, A., Eden, M., "The L_2 Polynomial Spline Pyramid," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 4, April 1993, pp. 364-379.

[56] Unser, M., Aldroubi, A., Eden, M., "A family of polynomial spline wavelet transforms," Signal Processing, Vol. 30, No. 2, January 1993, pp. 141-162.

[57] Unser, M., Aldroubi, A., Schiff, S.J., "Fast Implementation of the Continuous Wavelet Transform with Integer Scales," to be published.

[58] Chui, C.K., Wang, J.-Z., "A Cardinal Spline Approach to Wavelets," *Proceedings of the American Mathematical Society*, Vol. 113, No. 3, November 1991, pp. 785-793.

[59] Chui, C.K., Wang, J.-Z., "On Compactly Supported Spline Wavelets and a Duality Principle," *Transactions of the American Mathematical Society*, Vol. 330, No. 2, April 1992, pp. 903-915.

[60] Burt, P.J., Adelson, E.H., "The Laplacian Pyramid as a Compact Image Code," *IEEE Transactions on Communications*, Vol. COM-31, No. 4, April 1983, pp. 532-540.

[61] Mallat, S., Hwang, W.L., "Singularity Detection and Processing with Wavelets," *IEEE Transactions on Information Theory*, Vol. 38, No. 2, March 1992, pp. 617-643.

[62] Mallat, S., "Zero-Crossings of a Wavelet Transform," *IEEE Transactions on Information Theory*, Vol. 37, No. 4, July 1991, pp. 1019-1033.

[63] Vetterli, M., "A Theory of Multirate Filter Banks," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-35, No. 3, March 1987, pp. 356-372.

[64] Nalwa, V.S., Binford, T.O., "On Detecting Edges," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-8, No. 6, November 1986, pp. 699-714.

[65] Canny, J., "A Computational Approach to Edge Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-8, No. 6, November 1986, pp. 679-698.

[66] Berman, Z., Baras, J.S., "More about Wavelet Maxima Representations," in *Proc. 1993 IEEE ICASSP*, pp. V-662 to V-665.

[67] Yang, X., Wang, K., S., S.A., "Auditory Representations of Acoustic Signals," *IEEE Transactions on Information Theory*, Vol. 38, No. 2, March 1992, pp. 824-839.

[68] Irino, T., Kawahara, H., "Signal Reconstruction from Modified Wavelet Transform - An Application to Auditory Signal Processing," in *Proc. 1992 IEEE ICASSP*, pp. I-85 to I-88.

[69] Sen, D., Irving, D.H., Holmes, W.H., "Use of an Auditory Model to Improve Speech Coders," in *Proc. 1993 IEEE ICASSP*, pp. II-411 to II-414.

[70] Atal, B.S., "Automatic Speaker Recognition Based on Pitch Contours," Ph.D. Thesis, Polytechnic Institute of Brooklyn, 1968.

[71] Hammerlin, G., Hoffmann, K., *Numerical Mathematics*, Springer-Verlag, New York, NY, 1989.

[72] Prenter, P.M., *Splines and Variational Methods*, John Wiley & Sons, 1975.

[73] Schultz, M.H., *Spline Analysis*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1973.

[74] Keys, R.G., "Cubic Convolution Interpolation for Digital Image Processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-29, No. 6, December 1981, pp. 1153-1160.

[75] Chen, T.C., deFigueiredo, R.J., "Two-Dimensional Interpolation by Generalized Spline Filters Based on Partial Differential Equation Image Models," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-33, No. 33, June 1985, pp. 631-642.

[76] Sankar, P.V., Ferrari, L.A., "Simple Algorithms and Architectures for B-Spline Interpolation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 10, No. 2, March 1988, pp. 271-276.

[77] Battle, G., "A Block Spin Construction of Ondelettes. Part I: Lemarie Functions," *Communications in Mathematical Physics*, 110, 1987, pp. 601-615.

[78] Strube, H.W., "Determination of the instant of glottal closure from the speech wave," *Journal of the Acoustical Society of America*, Vol. 56, No. 5, November 1974, pp. 1625-1629.

[79] Noll, A.M., "Cepstrum Pitch Determination," *Journal of the Acoustical Society of America*, Vol. 41, No. 2, 1967, pp. 293-309.

[80] Harris, J.D., Nelson, D., "Glottal Pulse Alignment in Voiced Speech for Pitch Determination," *ICASSP 1993 Proceedings*, Vol. II, pp. 519-522.

[81] Krubsack, D.A., Niederjohn, R.J., "An Autocorrelation Pitch Detector and Voicing Decision with Confidence Measures Developed for Noise-Corrupted Speech," *IEEE Transactions on Signal Processing*, Vol. 39, No. 2, February 1991, pp. 319-329.

[82] McAulay, R.J., Quatieri, T.F., "Pitch Estimation and Voicing Detection Based on a Sinusoidal Speech Model," *ICASSP 1990 Proceedings*, pp. 249-252.

[83] Sondhi, M.M., "New Methods of Pitch Extraction," *IEEE Transactions on Audio and Electroacoustics*, Vol. AU-16, No. 2, June 1968, pp. 262-266.

[84] Kadambe, S., Boudreaux-Bartels, G.F., "A Comparison of a Wavelet Functions for Pitch Detection of Speech Signals," *ICASSP 1991 Proceedings*, pp. 449-452.

[85] Kadamber, S., Boudreaux-Bartels, G.F., "Application of the Wavelet Transform for Pitch Detection of Speech Signals," *IEEE Transactions on Information Theory*, Vol. 38, No. 2, March 1992, pp. 917-924.

[86] Fletcher, H., *Speech and Hearing in Communication*, Krieger Publishing Co. Inc., Huntington, New York, 1953.

[87] Fant, G., *Speech Sounds and Features*, MIT Press, Cambridge Mass., 1973.

[88] Flanagan, J.L., *Speech Analysis Synthesis and Perception*, Springer-Verlag, Berlin-Heidelberg, 1965.

[89] Hermes, D.J., "Pitch Analysis," in [16], pp. 3-25.

[90] Pratt, W.K., *Digital Image Processing*, Wiley & Sons, Inc., 1978.

[91] Shilov, G.Y., *Mathematical Analysis - A Special Course*, Pergamon Press, 1965.

[92] Natanson, I.P., *Theory of Functions of a Real Variable*, Volume I, Frederick Ungar Publishing Co., New York, 1961.

[93] Natanson, I.P., *Theory of Functions of a Real Variable*, Volume II, Frederick Ungar Publishing Co., New York, 1960.

[94] Furui, S., *Digital Speech Processing, Synthesis, and Recognition*, Marcel Dekker, Inc., New York, 1989.